LIBRARY

+8 173

Vol. XIX, No. 1 *PSYCHOLOGICAL REVIEW PUBLICATIONS* February, 1936

FEB 4 1936

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA

WALTER S. HUNTER, CLARK UNIVERSITY (*Psychol. Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Psychol. Rev.*)
JOHN A. McGEOCH, WESLEYAN UNIVERSITY (*Psychol. Bull.*)
JOHN F. DASHIELL, UNIV. OF NORTH CAROLINA (*Psychol. Monog.*

CONTENTS

Studies in Retroactive Inhibition: VI. The Influence of the Relative Serial Positions of Interpolated Synonyms: J. A. McGEOCH AND G. O. McGEOCH 1

The Relationship Between Speed and Accuracy in a Motor Task: B. R. PHILIP 24

Motor Concomitants of the Association Reaction: H. E. BURTT 51

Handedness Effect in Simultaneous Lifting of Weights by Both Hands: N. C. SHEN 64

The Conditioned Reflex as the Explanation of Habit Formation: III. The Operation of Two Higher-Order Reactions in Close Succession: J. M. STEPHENS 77

Changes in Neuro-Muscular Tension Accompanying the Performance of a Learning Problem Involving Constant Choice Time: E. GHISELLI 91

The Effect of Art Training on Mirror Drawing: C. D. FLORY 99

Certain Somatic Activities in Relation to Successful and Unsuccessful Problem Solving: M. S. CLITES 106

Discussion:

A Note on the Relative Efficacy of a Bell Signal for Right and Wrong Responses in Maze Learning: K. F. MUENZINGER 116

A Comparison of the Bell Signal for Right and Wrong Responses: A Reply: W. S. HULIN AND D. KATZ 117

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS

OF THE

AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

S. W. FERNBERGER, University of Pennsylvania (*J. Exper. Psych.*)
WALTER S. HUNTER, Clark University (*Index and Abstracts*)
HENRY T. MOORE, Skidmore College (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, Princeton University (*Review*)
JOHN A. McGEOCH, Wesleyan University (*Bulletin*)
JOHN F. DASHIELL, Univ. of North Carolina (*Monographs*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the Bulletin consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 900 pages.

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The Index is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consists of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The Monographs appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, June, September, December, March, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: $5.50 (Foreign, $5.75). **Index:** $4.00 per volume.
Journal: $7.00 (Foreign, $7.25). **Monographs:** $6.00 per volume (Foreign, $6.30).
Bulletin: $6.00 (Foreign, $6.25). **Abstracts:** $6.00 (Foreign, $6.25).
Abnormal and Social: $5.00 (Foreign, $5.25). Single copies $1.50.
Current numbers: **Journal,** $1.25; **Review,** $1.00; **Abstracts,** 75c; **Bulletin,** 60c.

COMBINATION RATES

Review and Bulletin: $10.00 (Foreign, $10.50).
Review and J. Exp.: $11.00 (Foreign, $11.50).
Bulletin and J. Exp.: $12.00 (Foreign, $12.50).
Review, Bulletin, and J. Exp.: $16.00 (Foreign, $16.75).
Review, Bulletin, J. Exp., and Index: $19.00 (Foreign, $19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, NEW JERSEY

Journal of
Experimental Psychology

VOL. XIX, NO. 1 FEBRUARY, 1936

STUDIES IN RETROACTIVE INHIBITION: VI. THE INFLUENCE OF THE RELATIVE SERIAL POSITIONS OF INTERPOLATED SYNONYMS [1]

BY JOHN A. McGEOCH AND GRACE O. McGEOCH

Wesleyan University

PROBLEM

It has been shown in a variety of ways that degree of retroactive inhibition is a function of the similarity of interpolated to original learning materials.[2] This relation has been interpreted in terms of a transfer theory of inhibition, but the statements of the theory have gone no farther than to say that transfer takes place and that the inhibition which the experiments measure is to be referred to this transfer. An adequate account of the phenomenon can be given only if we know much more than we do at present about the character and the conditions of the transfer. This knowledge would be equally important for an understanding of positive transfer.

[1] The preceding studies in this series are: J. A. McGEOCH, Studies in retroactive inhibition: I. The temporal course of the inhibitory effects of interpolated learning, *J. Gener. Psychol.*, 1933, **9**, 24–42; II. Relationships between temporal point of interpolation, length of interval, and amount of retroactive inhibition, *J. Gener. Psychol.*, 1933, **9**, 44–56; J. A. McGEOCH and F. McKINNEY, III. Retroactive inhibition in the learning of poetry, *Amer. J. Psychol.*, 1934, **46**, 19–33; J. A. McGEOCH and M. E. NOLEN, IV. Temporal point of interpolation and degree of inhibition, *J. Comp. Psychol.*, 1933, **15**, 407–417; and J. A. McGEOCH and F. McKINNEY, V. The susceptibility of prose to retroactive inhibition, *Amer. J. Psychol.*, 1934, **46**, 429–436.

[2] The major papers on the problem are cited in J. A. McGEOCH, The psychology of human learning: a bibliography, *Psychol. Bull.*, 1933, **30**, 1–62, and Learning and retention of verbal materials, *Psychol. Bull.*, 1934, **31**, 381–407.

The present experiments have been designed to discover the influence upon retroactive inhibition of one specific variable in the similarity-transfer-retroaction context.

These experiments depart from an earlier experiment by McGeoch and McDonald [3] in which it was found that synonyms were more productive of inhibition than were lists of antonyms, unrelated adjectives, nonsense syllables, or 3-place numbers. The four latter were accompanied by decreasing amounts of inhibition in the order named. When, moreover, the synonyms were classified in three groups with respect to their degree of synonymity to the adjectives in the original list, the three groups yielded retroaction in amounts which varied directly with degree of synonymity. The problem of how the interpolated synonyms operate to inhibit the recall and relearning of the original list remains and the experiments to be described test one possibility.

The interpolated synonyms of the McGeoch and McDonald experiment each occupied the same serial position as did the corresponding adjective in the original list. The hypothesis is plausible that the common factor of serial position may have been one important condition of their inhibitory effectiveness. Any general interpretation of learning which regards the memorizing of a list of items as a simple serial chaining of responses and which views inter-list effects as being highly specific in character would lead one to expect the greatest inhibitory consequences from similar interpolated items when the serial position of each is identical with that of its analogue in the original list. A decrease in inhibition with increasing disparity in the relative serial positions of the similar items in the two lists would also appear to be a reasonable expectation. An experimental test of the influence of relative serial position upon degree of retroaction will constitute not only a measurement of one condition potentially involved in transfer but also of the interpretation of memorizing as simple associative chaining. These experiments have duplicated the experimental conditions of the former one, but

[3] J. A. McGeoch and W. T. McDonald, Meaningful relation and retroactive inhibition, *Amer. J. Psychol.*, 1931, 43, 579–588.

have sought to determine whether variations occur in degree of retroactive inhibition with shifts in relative serial position. The relative position here meant is *relative inter-list serial position*, that is, the relative positions of similar items in the two lists, not relative position within a single list.

Experiment I [4]

Procedure

Lists of 10 two-syllable adjectives, plus an initial word which served as a cue and was not learned, were exposed at a 1.8-sec. rate on an electrically driven drum. All original lists were given 5 presentations, followed by a 10-min. interval and then by relearning to 3 successive perfect repetitions. The 6 conditions of the experiment represent as many variations in the character of the activity engaged in by the subject during the interval. Other conditions are constant. Under the *rest condition*, which serves as the standard or control in experiments upon retroaction, the subjects read jokes throughout the interval under instruction to mark the three best on each page. Under the five others, all of which are *work conditions*, the subjects were given 10 presentations of a second or interpolated list, followed by reading jokes for the remainder of the interval. This interpolated learning began 30 sec. after the completion of the fifth repetition of the original list. The 5 work conditions vary from each other only in the character or in the arrangement of the interpolated lists. The interpolations are described in the following enumeration. (1) Each adjective in the interpolated list was a synonym of the adjective in the corresponding position in the original list. (2) Each adjective in the interpolated list was a synonym of an adjective in the original list but appeared in a position two places removed in the forward direction from the position of that adjective in the original list. Thus the synonym of the adjective in serial position 1 of the original list appeared in position 3 of the interpolated list. The synonyms of the two last words in the original list occupied the two first positions in the interpolated series. (3) This condition is identical with (2) except that here the interpolated synonyms were 4 positions removed from their corresponding adjectives in the original list. The synonyms of the four last words in the original list occupied the first positions in the interpolated list. (4) Each adjective in the interpolated list was a synonym of one in the original list, but the interpolated synonyms were arranged in the reverse serial order of their corresponding adjectives. (5) The interpolated list was composed of adjectives unrelated in meaning to those of the original list.

The placing of the synonyms in the interpolated lists under Work conditions 2 and 3 requires discussion. The method used can scarcely be avoided if one is to use in the interpolated list synonyms of all of the adjectives in the original list and to displace the synonyms by any constant number of positions over at least a part of the list. Its implications should, however, be understood. It is simple and unequivocal to move each of the first eight synonyms two positions forward in relation to the corresponding adjectives in the original list, as in Work 2, but the synonyms of the last two adjectives

[4] This experiment has been assisted by a Grant-in-Aid from the National Research Council. We are indebted, also, to Dr. Arthur W. Melton of the University of Missouri for numerous critical suggestions concerning both the experimental method and the interpretation of the data.

cannot be moved forward in the same way. Likewise, as in Work 3, the first six synonyms can each be moved four positions forward but the last four then remain unplaced. The procedure followed, as stated above, was to place the last two (or four) words at the beginning of the interpolated list. In Work 2, for example, the synonyms of items 9 and 10 were, thus, placed in positions 1 and 2.

This procedure is unequivocal if one assumes that a list is learned as a circular series, but this assumption cannot be made. On the contrary, the subjects were instructed concerning the number of trials to be given and the instruction would lead them to treat each trial as a repetition of a unilinear list. This would be strongly reinforced by the fact that, as is usual in the anticipation method, each list was headed by an adjective which served as a cue but was not learned, that is, was not to be anticipated on each succeeding trial when the last word of the list appeared. Instead, when the last word had been in the window of the screen for 1.8 sec., the drum was reset with the cue word in the window as a signal for the beginning of another trial. The clear break thus introduced between trials would emphasize further the unilinearity of the series. The unilinearity is made still more definite by the fact that in moving the synonyms of the last two (or four) adjectives of the original list to the beginning of the interpolated list the cue word was skipped. The conditions, thus, should give the subjects a set to react to the items as a repeated unilinear series.

This method of placing the synonyms is still unequivocal on the assumption of unilinearity of series for the subject, but it does not permit us to say how much relative disparity has been introduced and thus to place the disparities under the different work conditions at known distances apart on a scale.[5] It can, however, scarcely be doubted that Work condition 3 has introduced more disparity than has Work condition 2, although we cannot say how much more. Work condition 4, in which the synonyms were arranged in the reverse order of their corresponding adjectives, may be said, likewise, to have introduced more disparity than have Work conditions 2 and 3. This is sufficient for the present problem.

The synonyms used in the interpolated lists had been selected from a large list by a group of 12 mature judges as being approximately equal in degree of synonymity and as being very closely related to their analogous adjectives. The unrelated lists were, likewise, composed of adjectives judged to be unrelated in meaning. Three different lists were used under the rest condition and three different pairs of lists under the last work condition (unrelated adjectives). Under the four other work conditions it is possible to arrange each list of interpolated synonyms in each of the four ways, *i.e.*, with each synonym in the same position as its adjective in the original list, two or four positions removed, or in a reverse order. Four pairs of lists, *i.e.*, four lists of adjectives and four of interpolated synonyms, were employed and in each case the interpolated lists were arranged in each of the four ways. The six conditions were gone through in a counterbalanced practice order. The lists for the rest condition and for the work condition with the interpolation of unrelated adjectives were rotated in

[5] It is, of course, possible to count the number of serial positions by which each synonym is removed from its corresponding adjective and to use this number as an index of the disparity. It is, however, of doubtful meaning when some of the shifts are in the backward direction as well as the forward within an unilinear list. When such a count is made, Work conditions 2, 3 and 4 represent mean relative disparities of 3.2, 4.8 and 5 positions, respectively. In Work conditions 2 and 3 the mean relative disparities are greater than 2 and 4, respectively, because of the displacement of the last two and four items backward to the beginning of the series.

groups of three and the four arrangements of lists under the other four work conditions were distributed to counterbalance differences in difficulty. The counterbalancing scheme gives to each condition equal average position; by using three different lists or pairs of lists under Rest and Work 5 the data are prevented from being functions of the difficulty of particular lists; by rotating the four lists and four arrangements over the four other work conditions differences in difficulty are cancelled out here, also; and thus all subjects are given equal opportunity for practice. Prior to the beginning of the experiment all subjects learned and relearned two practice lists to three successive perfect repetitions.

Thirty-six subjects, 20 men and 16 women, went through each of the six conditions described.[6] All subjects were, prior to the learning of the first practice list, unfamiliar with the anticipation method. All were university students.

The amounts learned during the 5 presentations of the original lists and during the 10 presentations of the interpolated lists have been measured in terms of the total numbers of correct anticipations. Retention has been measured in terms of four different scores: (1) Recall 1, which is the number of correct anticipations on the first relearning trial; (2) Recall 2, which is the number of correct anticipations on the second relearning trial and which serves as a useful supplement to Recall 1, particularly on account of the fact that it is less likely to be influenced by chance factors than is Recall 1; (3) Relearning 1, which is the number of trials required to relearn to the first perfect trial and (4) Relearning 2, which is the number of trials required to relearn to a criterion of three successive perfect trials. The number of presentations during the learning period is constant and it is, therefore, unnecessary to compute saving scores.

Results.—All of the original lists were given five learning presentations, but it is desirable to check this constant frequency by other measures in order to ascertain whether the degrees of original learning under the six conditions were the same. The mean number of correct anticipations per subject offers an additional measure by which frequency may be checked. The numbers per subject at the six conditions are: Rest, 18.50 ± 1.16; Work 1, 17.91 ± 1.11; Work 2, 17.86 ± 1.23; Work 3, 16.97 ± 0.97; Work 4, 18.66 ± 1.16; and Work 5, 19.69 ± 1.26. The differences between the means are small, of no statistical importance, and run in no regular order. It may fairly be assumed that any differences in retention which may be found are not importantly influenced by differences in measured degree of learning. The corresponding numbers for the five interpolated lists under the work conditions are 48.08 ± 2.73, 47.91 ± 2.48, 48.50 ± 2.34, 48.30 ± 2.47 and 54.50 ± 2.85, respectively. The first four

[6] The experiment was conducted by an assistant, Mr. Willis H. McCann, who had had a year's careful training in this kind of experimentation and whose work was under careful supervision throughout.

means are close together in magnitude and the deviation of the fifth is neither statistically reliable nor, in view of the fact that the unit of measurement is a single anticipation, very great.

The means of the recall and relearning scores, with their standard errors, are presented in Table I. It is unnecessary

TABLE I

MEAN RECALL AND RELEARNING SCORES

Interpolation	Recall 1	Recall 2	Relearning 1	Relearning 2 [7]
Rest	4.75±.48	6.68±.39	5.46±.84	9.00±1.38
Work 1 (Syn. Same Pos.)	0.72±.19	3.52±.34	8.72±.84	12.00±1.27
Work 2 (Syn. 2 Pos. Rem.)	0.44±.17	3.41±.40	9.63±.95	13.16±1.29
Work 3 (Syn. 4 Pos. Rem.)	0.44±.31	3.47±.33	9.38±.92	14.22±1.27
Work 4 (Syn. Reversed)	0.91±.28	4.30±.41	8.80±.77	12.00±1.26
Work 5 (Unrel. Adj.)	1.16±.30	5.38±.34	6.41±.61	9.63± .98

to translate the data into percentages of retroactive inhibition since the degrees of original learning are not importantly different and since the conditions under which the learning was done are the same. Each work condition is to be compared with the standard or rest condition.

The first thing which is apparent from an examination of the table is that each interpolation has produced at least some decrement in retention by every one of the four measures. The decrement is greatest in terms of Recall 1 and least in terms of the relearning trials. It is also clear that the interpolation of unrelated adjectives, while it is productive of a considerable amount of inhibition, produces less than does any of the other interpolations.[8] These results corroborate those of the preceding experiment and serve as assurance that the other results of the present experiment may be interpreted to be continuous with those of the prior one.

The important thing for the present problem is the rela-

[7] The mean relearning trials in these experiments do not include the perfect or criterial trials.

[8] The differences between the means of Work 5 and of the other work conditions are not statistically reliable, but the trustworthiness of their direction is strongly supported by the fact that, when the data are fractionated into three separate groups, including 12 subjects each, the three are consistent in direction.

tionship between variations in retroaction and changes in relative serial position. It is evident at once that, although the disparity between the serial positions of the adjectives in the original list and their synonyms in the interpolated list becomes greater from Work 1 to Work 4, there is no accompanying uniform variation in retention by any of the four measures. Identity of position is not an essential condition for the inhibitory action of interpolated synonyms, nor is this action a regular function of positional disparity.

There are, however, certain differences between the means of the first four work conditions which must be considered. Retroactive inhibition is greatest throughout at the Work 2 and Work 3 conditions. Between these two conditions there is no significant difference and we may treat them as indistinguishable from each other for the purposes of comparison with Work 1 and Work 4. It is doubtful if the differences obtained between these two conditions and Work 1 and 4 can be given much weight. None of them is statistically reliable [9] and their consistency of direction would weigh more heavily in favor of credibility if fractionation of the data also yielded consistent results. When, however, the data from the 36 subjects are divided into three groups of 12 subjects each and the means computed for each group separately, only two of the three groups show this consistency of direction. The means of the third group show one reversal in terms of each one of the 4 measures. The amounts of these reversals are not large but they are enough to be inconsistent with the two other groups. The differences between Work 1 and Work 4 are neither large nor consistent in direction.

The results are summarized in the following statements. (1) The interpolated learning of lists of synonyms consistently produces amounts of retroactive inhibition which vary with the method of measuring retention and which are consistently, although without statistical reliability, more than those produced by lists of adjectives unrelated to those of the original

[9] The largest single difference, that between the means for Relearning 2 at Work 1 or 4 and Work 3, which is 2.22, has a σ of 1.79. The sigmas of the other differences can be seen by inspection to be large, as large in most cases as the difference itself.

lists. (2) Interpolated synonyms which occupy the same position as, which are two or four positions removed from, and which are in the reverse order of their analogues in the original list do not vary among themselves by statistically reliable amounts with respect to the inhibition produced whether retention is measured by recall or by relearning.[10] There is a slight tendency, supported by all measures of retention in two of the three sub-groups which result from fractionation and in the total of all, for a disparity of two and of four positions to yield more retroaction than a disparity of none and than reversal. Since, however, this is small and is not supported in Experiment II, when a disparity of two and four positions is compared with a disparity of none, it may be considered as negligible.

EXPERIMENT II [11]

Procedure

The arrangement of the lists of interpolated synonyms in Experiment I leaves uncontrolled a variable which may be of importance. The relative disparities between the positions of the interpolated synonyms and those of their corresponding adjectives in the original list were produced by shifting all of the items in a uniform manner. In Work 2, for example, the synonym of item 1 in the original list appeared in position 3 of the interpolated list, the synonym of item 2 appeared in position 4 of the interpolated list, and so on. This brings it about that, under Work 2 and 3, the intra-list order of the synonyms is the same as that of their corresponding adjectives in the original list, except in the case of the last 2 and 4 synonyms, respectively, which were shifted back to the 2 or 4 initial positions. In Work 4, where the synonyms are in the reverse order of the original list, the intra-list order is a constant except for the reversal. It may be that the constancy of intra-list order in these conditions, and especially in Work 2 and 3, may have affected the results. The subject might, for example, think of some or all of the synonyms as related to the corresponding adjectives in the original list, thus in effect moving the list around circularly to an equality of relative position with the original list. This could account, if it occurred, for the failure of the presentational shifts in relative position to produce differing amounts of inhibition. The constancy of intra-list order could, conceivably, have had other effects, also, which would contribute to the obtained result. Experiment II has been done as a check upon the influence of this variable.

Experiment II duplicates Experiment I except for the substitution of two random orders of the interpolated synonyms in place of the reversed synonyms and of the unre-

[10] This conclusion is supported, also, by the mean numbers of correct anticipations per trial in relearning up to but not including the criterial trials, which may be taken as an additional measure of retention. The means are: Rest, 7.38; Work 1, 6.89; Work 2, 6.53; Work 3, 6.43; Work 4, 6.64; and Work 5, 6.58.

[11] This experiment was conducted by Dr. Grace O. McGeoch.

lated adjectives of Work 4 and 5. The Rest condition and Work conditions 1, 2 and 3 are the same as the corresponding conditions of Experiment I. In Work 4 the synonyms of the original list are interpolated in the random order 8–1–5–2–9–7–4–6–10–3 and in Work 5 they are interpolated in a second random order, namely, 9–5–1–10–6–3–8–4–7–2. These two random orders break up the intra-list sequence and provide a check upon intra-list sequence as a condition of the results.

Eighteen girls, students at Stephens College, went through each of the six conditions in a counterbalanced practice order similar to that used in Experiment I. All subjects were, prior to the learning of the first of the two practice lists, unfamiliar with the anticipation method and with any experimental work on learning.

Results.—The mean numbers of correct anticipations per subject during the five [12] learning presentations are: Rest, 16.28 ± 1.35; Work 1, 16.72 ± 1.42; Work 2, 16.39 ± 1.16; Work 3, 16.28 ± 1.32; Work 4, 15.61 ± 1.11; and Work 5, 17.28 ± 1.39. The differences between the means are small, statistically unreliable and irregular in direction. There is no reason to suppose that whatever differences may appear in retention will have been influenced by these small differences in degree of learning. The mean numbers of correct anticipations per subject during the 10 repetitions of the interpolated lists are: Work 1, 46.28 ± 3.60; Work 2, 46.28 ± 2.55; Work 3, 45.00 ± 2.83; Work 4, 52.17 ± 3.61; and Work 5, 47.22 ± 2.69. The mean under Work 4 deviates somewhat from the means of the four other work conditions, but by statistically unreliable amounts. The fact that it occurs with one of the interpolations in a random sequence and that the other has a mean which does not deviate greatly from the means of the other conditions permits a check upon its influence. There is in the measures of retention no indication that it has had any important effect. The means for both original and

[12] This relatively small number of learning presentations, which is insufficient for any complete mastery of a list, was used for two reasons: (1) to make the results of these experiments comparable with those obtained by McGeoch and McDonald and (2) because it is probable that associations with serial position will have a greater opportunity to show an inhibitory influence when learning is incomplete. In this experiment the mean numbers of correct anticipations at the last trial in original learning were 5.6, 5.3, 4.9, 5.6, 5.3 and 5.9 for the six conditions respectively. The corresponding numbers in Experiment I were 6.5, 5.6, 6.0, 5.7, 6.2 and 6.5. Serial position curves have been plotted from the data upon both original and interpolated learning in the two experiments and have been found to have the usual form. There is no reason to believe that the incomplete learning could have spuriously affected the results obtained and, furthermore, the serial position curves for retroactive inhibition give no evidence of such an effect.

interpolated lists are, for the most part, slightly smaller than the corresponding means in Experiment I. The somewhat greater age of the subjects in Experiment I could easily account for this. The relatively high agreement between the results obtained by two different experimenters upon samples from two very different institutions is more important than the small divergences mentioned and is one guarantee of the constancy of the experimental conditions.

The means of the recall and relearning scores, with their standard errors, are given in Table II. For the reasons already given in connection with Table I there is no need to express the results in terms of percentages of inhibition.

TABLE II

MEAN RECALL AND RELEARNING SCORES

Interpolation	Recall 1	Recall 2	Relearning 1	Relearning 2
Rest	4.50±.63	6.50±.50	5.55±.76	9.05±1.25
Work 1 (Syn. Same Pos.)	.72±.21	3.00±.55	8.44±.87	12.17±1.46
Work 2 (Syn. 2 Pos. Rem.)	.22±.09	2.11±.35	9.17±.89	11.72±1.29
Work 3 (Syn. 4 Pos. Rem.)	.55±.18	3.00±.49	7.66±.89	10.83±1.25
Work 4 (Syn. Random Ord. *x*)	.61±.18	3.39±.41	8.55±.82	11.88±1.48
Work 5 (Syn. Random Ord. *y*)	.11±.10	2.88±.51	8.39±.95	10.50±1.42

The mean recall scores under the rest condition differ only very slightly from the corresponding scores in Experiment I, while the relearning scores under rest in the two experiments are almost identical. This close agreement between the two experiments is a further indication of the constancy of the experimental conditions. The means of the comparable work conditions of the two experiments differ somewhat more, but the differences are seldom large and are in no case statistically reliable. Every interpolation has produced a decrement in retention by every one of the four measures. Again it is greatest in terms of Recall 1 and least in terms of the relearning trials.

We shall ask first whether the results of Work conditions 1, 2 and 3 agree with those of the analogous conditions in Experiment I. The means for Recall 1 are smaller at Work 2 and 3 than at Work 1, but they do not decrease as disparity

of relative position increases. The differences involved are not statistically reliable and are not, on account of the large number of subjects giving zero scores, to be given any great weight. The means for Recall 2 are identical under Work 1 and 3 and are larger by unreliable amounts than the mean under Work 2. When retention is measured by the number of trials to relearn to one perfect repetition (Relearning 1), the most rapid relearning occurs under Work 3 and the least rapid under Work 2, and when it is measured by the trials required to reach a criterion of three perfect in succession Work 3 again shows the most rapid relearning while Work 1 shows least. Work 3 yields, thus, a greater recall and a more rapid relearning than Work 2 and a more rapid relearning than Work 1. None of the differences involved is reliable statistically. The added fact that the direction of the differences frequently reverses that of the analogous differences in the data of Experiment I, can only be taken to mean that no important differences have been demonstrated in the inhibitory potencies of the three interpolations.

The next question and the one which Experiment II was especially designed to answer is whether differences exist between the inhibition produced by a uniform shifting of the relative positions of the interpolated synonyms and that produced by a random arrangement. An answer is to be sought by comparing retention scores under Work conditions 4 and 5 with those under Work conditions 1, 2 and 3. The two random orders themselves have yielded slightly different amounts of inhibition. Recall scores are unreliably higher under the *x*-order than under the *y*, while relearning is unreliably more rapid under the *y*-order, and it may be concluded that no important difference between them has been found.[13] The results from the interpolation of the two random orders give an indication of the range of retention which may be expected from the interpolation of such orders. The differences in question are of the same order as or slightly smaller than those between Work conditions 1, 2 and 3.

[13] The difference in terms of Recall 1 deserves little weight on account of the general unreliability of Recall 1 scores under the work conditions.

When Work conditions 4 and 5 are compared with Work conditions 1 to 3, no reliable or consistent differences are found. Both recall and relearning scores under the *x*-order are very nearly the same as the corresponding scores under Work 1 and Work 3 and differ only slightly more from those under Work 2. The same statement holds, with minor and unreliable exceptions, for the *y*-order. Variations occur in the retention scores under the five work conditions but they occur without consistency and in statistically unreliable amounts. Discussion of the individual variations from condition to condition would be fruitless.

The data of Experiment II lead to the following conclusions. (1) Interpolated synonyms which are arranged in a random order with reference to their corresponding adjectives in the original list produce approximately as much inhibition as synonyms which appear in the same order as their corresponding adjectives or which are two or four positions removed from this order.[14] (2) This experiment agrees with Experiment I in finding no important differences in the amounts of inhibition yielded by interpolation of synonyms in the same order as, and in two or four positions removed from, the corresponding adjectives in the original list. (3) The failure of the two experiments to find such differences, when interpreted in the light of the equivalent results produced by both uniform and random shifts in relative position, is not a function of the uniform shifting of relative serial positions.

Serial Position Effects in the Two Experiments

Serial Position Effects in Learning.—It is desirable, for an interpretation of the results, to know whether the subjects adjusted their learning of the interpolated lists to correspond in any measured way to the changes in method of arrangement. They might, for example, have perceived the relations between the original and interpolated lists and the method of arrangement and might have used the already practiced original list as an aid in learning the synonyms. If this had

[14] These conclusions are supported by the data for mean anticipations per trial during relearning up to but not including the criterial trials. The means are: Rest, 7.32; Work 1, 6.15; Work 2, 5.89; Work 3, 6.28; Work 4, 6.28; and Work 5, 5.81.

been done, it might show itself in some alteration of the serial position curve for the interpolated lists. The first two positions in Work 2 and the first four in Work 3 might show a relative disadvantage because of the departure of the synonyms at these positions from the uniform forward shift over the remaining positions. Whatever may be our hypotheses about possibilities, it is desirable to know whether any differences appear between the serial position curves of the different interpolations.

We have found the mean numbers of correct anticipations per subject per trial for both the original and interpolated lists in each experiment and have systematically compared the values under the different conditions. The serial position curves for interpolated learning have, under each condition, the same form as do those for original learning. The curves for the various arrangements of the interpolated synonyms in Experiment I show no regular variations in form and have the same form as does the curve for unrelated adjectives. The curves for the first three work conditions of Experiment II do not, likewise, differ significantly from each other or from those of Work 4 and 5 in which the interpolated synonyms appeared in a random order. In both experiments the rank-difference correlations between the serial position values under Work 1 and under the other work conditions vary from 0.77 to 0.98. The coefficients are, moreover, often lowered by slight shifts from one condition to another which are of no significance but which are sufficient to affect the rankings. It can only be concluded that the positions of the interpolated synonyms with reference to the positions of their analogues in the original list has had no measured influence on their serial position values.

Serial Position Effects in Recall and Relearning.—It is possible that the different arrangements of the interpolated lists may have produced different serial position effects in the recall of the original lists. It might be, for example, that the shifting of the synonyms of items 9 and 10 in the original list to positions 1 and 2 of the interpolated list would give at these two positions measurable differences in degree of retroaction

from the values at the same positions when the synonyms appeared in the same order as that of the original list. Or it might be that the serial position values under the conditions of uniform shift in relative position would differ from those under interpolation of unrelated adjectives or of synonyms in a random order. Serial position values have been found in terms of (1) Recall 1, (2) Recall 2, and (3) Trials to Relearn to the point beyond which no error is made. Percentage of retroactive inhibition has been found in terms of each of these scores at each position by subtracting the mean for each position under work from that under rest and dividing by the mean under rest.

The resulting percentages of inhibition have been examined in detail. The only justifiable conclusion is that there are no clear differences between the different work conditions with respect to the forms of the serial position curves. The percentages in Work 5 of Experiment I, in which unrelated adjectives were interpolated, are smaller in amount than those under the other work conditions, but they yield a curve of the same form. Work conditions 4 and 5 of Experiment II, in which synonyms were interpolated in a random order, likewise give serial position values which do not differ regularly from those for interpolated synonyms in either the same or a progressively displaced order. Occasional suggestions of a general tendency appear in one experiment only to be contradicted by an opposite tendency in the other. Since the experiments agree well in other respects, we may take these contradictions as an indication of 'chance' intraserial effects.

There is nothing to be gained by presenting tables of all of the percentages of inhibition. Two sample tables will be given. Both are for Experiment II because that experiment includes the two random orders which are of particular importance as a check upon the results from the more regular orders. The data for Recall 1 are unstable, although they agree with those obtained by the other methods of scoring, and we shall give the results for Recall 2 and Relearning 2 which are more stable. Any percentages of inhibition computed for each item from this number of records are, however,

relatively unstable and a clear tendency must be present before any positive conclusion can be drawn.

The percentages for Recall 2 appear in Table III. The

TABLE III

PERCENTAGES OF INHIBITION AT EACH SERIAL POSITION

Recall 2

Condition	Serial Position									
	1	2	3	4	5	6	7	8	9	10
Work 1 (Syn. Same Pos.)	41.2	64.7	39.9	63.6	54.6	80.0	57.2	42.8	37.6	45.7
Work 2 (Syn. 2 Pos. Rem.)	41.2	70.6	73.3	90.9	100.0	80.0	57.2	71.4	75.0	18.2
Work 3 (Syn. 4 Pos. Rem.)	28.4	47.0	53.4	72.8	81.8	88.1	42.8	57.2	50.0	36.5
Work 4 (Syn. Random Order *x*)	35.2	52.9	39.9	63.6	81.8	70.0	28.6	85.4	12.6	00.0
Work 5 (Syn. Random Order *y*)	17.6	41.2	80.0	63.6	81.8	90.0	42.7	71.3	62.6	18.2

serial position curves for retroactive inhibition do not have the form of the analogous curves for number of correct anticipations in learning or relearning. They are more irregular and the initial and final positions have a less clear advantage. There is in the data of Table III no large deviation at any point under one condition which is not deprived of its significance for that specific condition by a similar deviation at the same point under a different condition and there is none which has a common implication in the two experiments.

The percentages for Relearning 2 (Table IV) have been found directly from the mean trials required at each position

TABLE IV

PERCENTAGES OF INHIBITION AT EACH SERIAL POSITION

Relearning 2

Condition	Serial Position									
	1	2	3	4	5	6	7	8	9	10
Work 1 (Syn. Same Pos.)	243.0	53.2	100.2	62.9	37.6	39.7	54.3	70.1	226.7	205.8
Work 2 (Syn. 2 Pos. Rem.)	317.1	34.9	61.6	45.7	37.8	35.2	29.3	54.2	206.8	168.2
Work 3 (Syn. 4 Pos. Rem.)	142.8	59.2	89.6	+03.7	+04.7	31.0	30.2	38.3	124.5	143.9
Work 4 (Syn. Random Order *x*)	211.4	81.9	72.9	29.4	14.4	37.1	31.9	55.3	110.5	105.8
Work 5 (Syn. Random Order *y*)	258.0	81.9	75.9	22.8	10.3	34.4	38.5	57.4	96.3	70.7

to reach the point beyond which no error was made. The inhibition is much greater at the initial and the two final positions than at the intermediate ones.[15] The two things which at once attract attention are the slight facilitation effects at positions 4 and 5 under Work 3 and the regular decline in inhibition from Work 1 to Work 5 at the two final positions. The facilitation effects, however, not only do not appear in Work 3 of Experiment I but the inhibition is relatively higher there at the corresponding positions than at adjacent ones. There is, likewise, no trend toward a decline from Work 1 to Work 3 in Experiment I. In the same way the other and less obvious features of the table can be analyzed to show that no clear serial position effects are correlated with variations in the relative positions of the related items in the original and interpolated lists.

The fact that the interpolation of synonyms in a random order presents the same picture, with respect to serial position values, as do the more regular orders, is important for an interpretation of the mutual interrelations of the original and interpolated lists. The interference effects must, in the light of these results, be treated in some other way than as a one-to-one interference of items with relative position as a factor and must take account of the freedom of the interference from any dependence on relative position.

Recognition of Synonymity

Pairing of Synonymous Adjectives.—The experimental conditions permit little doubt that the synonymity of the items in the original and interpolated lists is a major variable. We have, however, wished to make sure that the pairs used were known as synonyms by the subjects who learned the lists, since absence of such knowledge would, presumably, represent a limitation upon the influence of synonymity. At the close of Experiment II the subjects were asked to pair the adjectives which had appeared in the interpolated lists with the corresponding adjectives which had been in the original lists. The

[15] The more general problem of the relation between serial position and percentage of inhibition is to be treated in another paper.

individual adjectives were arranged in two sets in a random order with respect to their relations with each other and to the lists in which they had been learned. Each set contained two lists of 25 numbered adjectives, one list being words from original lists and the other their synonyms in the interpolated lists. The subjects were asked to place opposite each word in the first list the number of its synonym, if any, in the second. They were not told that every adjective in one list had a synonym in the other, but only that they were to select every word in the second list which seemed to them a synonym of one in the first.

Out of the 50 possible pairings by 18 subjects, or 900 pairings in all, there were 30 which actually were not paired by the subjects or which were paired incorrectly. This is an average of 1.66 per subject and is a very small percentage (3.3) of the 50 pairs learned by each. Twenty-five of the failures were, moreover, made by four subjects. The records of these subjects have been examined for evidence of any differential susceptibility to inhibition, but none has been found.

It may be concluded that the synonymity used was recognizable to so high a degree that it may validly be considered the major variable it was designed to be. It is a fair assumption that the results in this respect from the subjects in Experiment II are representative of those which would be found with the subjects of Experiment I.

Recognition of Synonymity During Learning.—The method first described has shown that the subjects could recognize as synonyms nearly all of the pairs used when they appeared in parallel columns on a single sheet of paper. It is an interesting question whether they recognized the synonymity when the words were presented serially and when the members of each pair were in different lists. An answer is not crucial to the present experiments, but we have attempted briefly to obtain relevant data. In a third experiment, six Stephens College girls went through the conditions of Experiment II in a counterbalanced order and were asked, at the completion of relearning under each work condition, whether they had observed any relation between the original and interpolated

lists, what the relation, if any, was, and what they thought its influence had been on retention.

Two of the six subjects recognized throughout the experiment that the lists were synonymous and had, under most conditions, some appreciation of the plan of arrangement. Two recognized some pairs as synonymous but did not observe that all of the interpolated adjectives were synonyms of those in the original list and were not aware of the plan of arrangement. Two recognized no relation between the two lists. The subjects' reports upon the influence of the interpolated learning on retention cast no light upon the present problem and may be omitted. The number of cases is much too small to warrant a conclusion, but as far as they go they show no relation between recognition of the relation and amount of inhibition.

Discussion

We have found in two experiments that the inhibition produced by the learning of interpolated synonyms does not vary with the position of the synonyms in relation to their analogues in the original list. This absence of correlation between degree of inhibition and disparity in relative position contradicts at least one expectation from interpretations of learning as being primarily the serial and self-contained chaining of isolated items. It would be reasonable on such a view to expect that interpolated synonyms should exercise their greatest inhibitory effects when each synonym occupies the same position as its corresponding adjective in the original list. A subject learns a list as a series of successive items, each of which has a certain position, either absolute or relative, and while the association of each item with its position is not complete such association occurs and often to a considerable degree. The transfer incident upon interpolated learning might be expected to be greater if the related items are in identical serial positions, since they are then connected by this identical element. When the factor of identical position is removed, the inhibitory effect of the synonyms might be expected to become more nearly that of unrelated adjectives

and it might be that this would be more true the greater becomes the disparity in relative position.

These expectations are not borne out by the data. The amount of inhibition does not vary regularly with changes in relative position and there is no tendency for the inhibition from interpolated synonyms, whatever their relative positions, to approach that from unrelated adjectives. Synonyms produce more inhibition than do unrelated adjectives no matter in what position the synonyms are placed with relation to their analogues in the original list.[16] The interpolations of synonyms, in whatever order, constitute a single and not greatly differentiated group of conditions, all of which are separable, in terms of inhibitory effects, from the interpolation of unrelated adjectives.

The immediate theoretical problem which the data present is to account for the fact that the inhibitory effects of the synonyms are independent of relative position. This problem is one phase of the more general one of the character and conditions of transfer. The fact that degree of inhibition is a function of the character of the materials learned and particularly of their similarity shows that transfer is at least a major general condition of the inhibition. Formulations of transfer theory have, however, done little more than state the region in which the conditions of inhibition are to be looked for, and we shall raise, first, some of the questions which a more adequate statement of transfer, as a theory of inhibition, should answer.

A transfer theory faces at least three general questions. (1) What is the direction of the transfer? Is it from original list to interpolated, from interpolated list to original, or in both directions?[17] (2) What is the temporal locus of the

[16] This may be shown by computing percentages of inhibition from the data of Tables I and II. The percentages for the unrelated adjectives of Experiment I are 75.5, 19.4, 17.3 and 07.0, for Recall 1 and 2, Relearning 1 and 2, respectively. The corresponding percentages for the two interpolations of synonyms in a random order in Experiment II are 86.6, 47.8, 54.0 and 31.2 for Order *x* and 97.5, 55.6, 51.1 and 16.0 for Order *y*.

[17] This question has been attacked experimentally in a preceding paper. *Cf.* F. McKinney and J. A. McGeoch, The character and extent of transfer in retroactive

transfer? Does it take place during the interpolated learning, during the interval between learning and relearning, or during relearning? [18] (3) In what way does the transfer occur and what are its basic conditions? Does it go on in terms of identical elements, generalization, equivalent stimuli, meaning, irradiation, assimilation, transposition of pattern, or of any combination of these or other interrelated possibilities? However it goes on, what are the conditions of its occurrence? This group of problems is the theoretically most important of the three. It is the one with which the present experiments are concerned.

We shall outline briefly in the light of the present data certain things which an attempt to answer these questions must consider and certain theoretical possibilities. An attempt to state a more general theory will not be made here. It is clear at the outset that, to the extent that the inhibition of one list by the other is not a function of relative serial position, there is a kind of total action of list on list, an action which is, however, specific to their general content since more similar materials yield more retroactive inhibition than do less similar. To say this merely states the problem in another way. We must look for some way in which one list may inhibit another without regard to relative position and as a function of the similarity between the items. We shall outline one possibility.

Within each list each word is connected as response to the preceding word and as stimulus to the succeeding one.[19]

inhibition: disparate serial lists, *Amer. J. Psychol.*, 1935, **47**, 409–423. Some transfer was found in both directions but it was much greater from interpolated list to recall of original list than from learning of original list to recall of interpolated. Under the conditions used inhibition was found to be a two-way interference phenomenon with one direction dominant.

[18] The data published by McKinney and McGeoch, *op. cit.*, imply that the temporal locus of the interference is at the time of recall. That it is during the measurement of retention is implied, also, by the data of J. A. McGeoch, Studies in retroactive inhibition: I. The temporal course of the inhibitory effects of interpolated learning, *J. Gener. Psychol.*, 1933, **9**, 24–42.

[19] To the extent that there are remote forward and backward associations each item is connected, also, with some or with all the other items in the list. We may expect remote associations to be roughly equal in original and interpolated lists and not to be a differential factor in this particular problem.

Each word, has, also, two general contexts: an external or environmental and an internal or representative context. It is, potentially at least, connected with the whole experimental situation, including the experimenter, the drum, the instructions, the furniture of the room, and other similar things. This environmental context is the same for every word in the list. There is, also, the subject's own organization, particularly his verbal and other ideational organization, including the complex meanings of each word in the list, meanings which each one may at least partially arouse. This constitutes a representative context which is specific to each item.[20]

Consider a word, as *perfect*, in the third position of an original list and its synonym, *faultless*, in the seventh (or any) position of the interpolated list. In the subject's vocabulary organization the two words have a common core of meaning regardless of their relative serial positions in these particular lists. They have, also, of course, certain individual connections specific to each. In the learning of the two lists each word has become associated with the common environmental context and at the same time, its already formed connections with the subject's representative organization have been aroused. When the stimulus to the recall of *perfect* is given, it comes in the common environmental context which is not specific to either list, and it rearouses the core of meaning common to it and to *faultless*, which in turn has also been connected with the non-specific environmental context. There is no reason to suppose that the rearousal of this common context is a matter of relative serial position, since serial position has no specificity to either.

The serial stimulus to the recall of *perfect* may, by reason of the meaning common to *perfect* and to *faultless* and not specific to any list or position, and by reason of their common environmental context, arouse both words either partially or completely. The result will be a blocking and failure to recall

[20] A separation of environmental and representative contexts can be made only arbitrarily. It is convenient to separate them here, since the context which is common to all words in each list is mainly the perceived experimental setting, while the context which varies from word to word but which is partially common to synonymous pairs is clearly representative.

correctly within the prescribed time or at all. Likewise, by reason of the meaning common to the two words, the serial stimulus to one may elicit the other instead or will lead, that is, to an incorrect recall in that particular list. Either word, again, when presented as stimulus, may arouse the response connected with the other word, *i.e.*, in the other list, as well as or instead of the one in its own list, again with inhibitory effect. It is probable that the common context of meaning already established in the subject's organization is more important than the common environmental context of the words.

This possibility assumes that either word of a synonymous pair may function in these contexts for the other. To what extent one will function for the other depends on the degree of their common meaning for the subject,[21] on degree of learning and other such factors. Presumably kind and amount of common meaning is the most important. The source of inhibitory potency lies, on this view, primarily in the relation of the material being learned to the already existent representative organization of the learner.

The items in a list are almost certainly associated to some degree with their serial arrangement. This association is probably seldom a point-to-point association of item and position; it may be areal only; but it is almost certainly present in some subjects, if not in all. Under some experimental conditions this association may be a factor in producing retroactive inhibition. In the present experiments there is no evidence that it has been effective. The hypothesis described attempts to picture conditions under which inhibition may occur independently of relative serial position. It is not, as it stands, a general theory of inhibition, nor is it a complete hypothetical account of the conditions of the present results. It is hoped that further experiments will permit a more adequate and more complete hypothesis.

[21] The hypothesis under discussion fits well the results of J. A. McGeoch and W. T. McDonald, *op. cit.* These results showed degree of retroaction to vary directly as degree of synonymy. This is what would be expected in terms of the present hypothesis. Other data upon the influence of similarity upon degree of inhibition are, also, fitted well by the hypothesis.

Conclusions

1. Ten learning presentations of a list of synonyms interpolated immediately after 5 presentations of an original list of adjectives have a pronounced inhibitory effect upon recall and relearning after an interval of 10 min. The inhibition produced by the synonyms is greater than that produced by a list of adjectives unrelated to the words of the original list.

2. No clear variation occurs in degree of inhibition when the synonyms in the interpolated list, instead of appearing in the same positions as their corresponding adjectives in the original list, are removed two or four positions, are presented in the reverse order, or are presented in a random order. *Inhibitory potency is, thus, independent of the relative positions of the synonymous adjectives in the two lists.* Serial position is, therefore, an unimportant factor in the production of inhibition under these conditions.

3. This result is difficult to interpret on a theory of memorizing as simple serial chaining and on a theory of transfer in terms of individual identical elements. One tentative hypothesis is suggested in terms of common contexts which become more and more completely common as the words approach identity of meaning and which constitute a general factor operating to produce interference between items of similar meaning. Serial position may be expected to have little specific influence on the operation of such a common factor.

(Manuscript received October 17, 1934)

THE RELATIONSHIP BETWEEN SPEED AND ACCURACY IN A MOTOR TASK

BY B. R. PHILIP

University of Alberta

Introduction

The study of the interrelations and variations of the speed factor in different types of activities is one of increasing interest in experimental psychology. One aspect of the general problem, that of the relationship between speed and accuracy, has been investigated only in a general way.

The results seem to indicate that there is a relationship between speed and accuracy so complex that correlational analysis of the performances of large groups of *S*s did not yield a satisfactory solution of the problem. Not much more satisfactory was the detailed analysis of the performances of few *S*s. Possibly the only definite finding was that each *S* worked at an optimum rate achieving a maximum speed with a minimum number of errors. As far as the author is aware, no attempt has been made to maintain one variable constant and systematically to trace the fluctuations of the other variable. That such a standard scientific procedure should be overlooked is quite surprising and is the probable reason for the lack of progress in solving the problem. The inadequacy of common statistical procedure may be partly ascribed to the complexity of the relationship between the variables. Then too the tasks chosen were not always satisfactory from the standpoint of range, reliability, and evaluation of scores.

It was decided to attack the problem using a method that uniformily is helpful. It was planned to hold the speed constant and to determine the accuracy at that particular speed. By varying the speed within a suitable range, all accuracies between 0 percent and 100 percent accuracy would be obtained and the full relationship determined. A suitable task

had to be chosen that would satisfy the requirements just mentioned. Certain refinements of technique were planned and a special self-registering apparatus was designed to facilitate the solution of the problem. A preliminary investigation with this object in view is reported in this article. The task chosen was a simple motor performance, that of continuous tapping.

Apparatus.—With the object of securing a self-registering portable instrument, adaptable to several different tasks, a special apparatus was designed and built. It is a modification of the McDougall tapping machine. Essentially, it consists of a perforated strip which revolves at variable rates. The tapping is done by means of a stylus, and the counting, both of successes and of errors, is automatic. Two strips of oilcloth, 56.25 inches in length, one 6 inches wide and the other 3 inches wide, are carried around 4 brass cylinders mounted on ballbearings. The rotation is accomplished by means of a small 100 volt A.C. motor, the speed of which is reduced by a spiral gear, and also by means of a pulley-drive. Thus reduced, the speed is varied by means of two rheostats so that the time of one revolution is from 5 to 45 seconds. The motor is mounted on sponge rubber and is firmly attached to a panel inside the revolving strips. Thus a smooth, powerful and fairly quiet revolution of the belts is obtained.

The two strips of oilcloth with their rough surface revolving over the cylinders each carry a different task. Task I, the complex series, consists of 25 perforations of one type (the *right* ones) which must be tapped as the strip appears in front of a window, and 12.5 (the *wrong* ones) which are not to be tapped. Thus at every 4 revolutions of the belt, 100 right and 50 wrong perforations appear. The right and the wrong perforations are mirror images of one another. They are rhomboids 0.25″ wide, 0.5″ long on the wide side and 0.25″ long on the narrow side. The rhomboid shape of the perforations was chosen so as to demand as keen a discrimination as possible when the strips are revolving, and at the same time to make the rights and the wrongs equal in area. The perforations occur at equal intervals of 1.5″ along the strip but are randomly distributed across the width of the strip except for this fact that the rights always come down the same channels and the wrongs down the others.

The strips slide between two metal walls, which keep them in a uniform path, on the rollers, which supply the traction, and over a metal plate on which the tapping is done through the perforations. A simple device permits of automatic registration of the rights and the wrongs on two electric counters. The circuit is from the brass stylus, through the perforations in the strips, on the metal plate, to the counters and is actuated by a 6 volt source which is between the counters and the stylus. If *S* misses the perforation and hits the oilcloth the circuit is not completed and no registration is made. As 100 perforations pass in one task, any misses can easily be calculated from the reading of the counter.

The differentiation of the rights and the wrongs is obtained as follows: The metal plate consists of two parts, which fit into each other but do not touch, and hence are insulated from each other. The fitting edges are denticulated. One part has 4 denticulations each approximately 0.5″ wide and 2″ long. If one imagines the total 6″ width of the strips to be divided into 12 equal channels, each 0.5″ in width, the denticulations occur at channels 3, 5, 8, and 11. The other part of the metal plate,

since it fits exactly into the former, has 5 denticulations, some of them of double width, and occurring in channels 1, 2, 4, 6, 7, 9, 10, and 12. The perforations in the oilcloth strip slide directly over these channels, the right ones always over the part of the metal plate with the more denticulations on it, and the wrong ones over the other part. These sections are connected respectively to the counters for rights and wrongs. As *S* looks through a window at the perforations on the strips as they slide by, no hint can be obtained as to the channeling of the rights and the wrongs. Even the knowledge of the way of the rights and wrongs come is of no help to *S* in the actual test.

The second oilcloth strip contains Task II. The perforations, 50 in number, are rectangular 0.5 " by 0.344" wide; they come down 4 channels in random order, and the channels are 0.5" wide. The interval between perforations is 1.125". There are no wrong perforations. The strip is 3" wide, and slides under a window between metal panels over a uniform brass plate which is connected to one of the mechanical counters. Contact is made by a stylus which is connected by a flexible cord to a 6 volt source, and through the counter to the metal plate: thus the circuit is similar to that of the first task. The strips revolve over the rollers and the speed is varied by the same arrangement of geared motor and variable rheostats. The strips are readily detachable by pulling out a metal pin which slides through a hinge-joint at the end of the strips. Hence it is a simple matter to adapt the apparatus to other tasks.

Task III is based upon Task I. Two rhomboid wells, of the same area and shape as the perforations of the first one, 0.25" in width, with 0.5" and 0.25" sides, are set in a brass plate, 0.125" below the surface of the plate. They are placed with the narrow sides innermost and at a distance of 2.75" apart, which is the average displacement of the perforations of the first strip. The wells are insulated from the plates and are connected to one of the electric counters. The stylus taps them alternately and the circuit is completed as in the other tasks.

The pair of styli are of special design. They are made of brass, are 6" in length and 0.375" wide. A plunger that makes the contact is 0.125" wide, 2" long, with a rounded point. It slides inside the brass sleeve, below a metal spring. A flexible connection is attached to the top. Thus an adequate contact is made without undue pressure.

The counters are standard Cenco impulse counters. They record accurately up to 120 impulses a second and the reading is taken directly from the dials which are mounted upright on the panel of the instrument. The electric supply is of two types, a 110 volt A.C. led directly from the mains, through a switch mounted on the panel, and a 6 volt D.C. supply which comes from an Elkon dry disk rectifier built in the instrument and mounted on a panel between the revolving strip. The rectification is excellent, and as tested on a General Electric oscillograph shows a straight line with ripples of very small amplitude. No filter nor choke is needed beyond the capacity and inductance of the circuit. The fact that discrete contacts are recorded on the delicate counters without any spinning or skipping of the dial hands warrants the inclusion of this type of D.C. current. Its advantages are of course evident. It is noiseless, light, never needs attention, though it has been in operation over two years, and is built directly in the instrument. A switch mounted on the panel sends the 110 volt A.C. into the rectifier. The 6 volt supply is for the counters and the lights.

The panel in front of the instrument is of bakelite. On it are mounted two colored lights, one green and one red; they are flashlight bulbs, shining through colored glass studs. Metal fingers make contact with two brass studs which are set in the revolving strip of the first task, and connect with the lamps. At each revolution of the strip,

first the green light flashes as a warning, then shortly after, depending upon the speed of revolution of the strip, the red light flashes as a signal for *E* to start the task. Thus he is enabled to count the number of revolutions of the strip, generally 4 to a task. As an aid in checking the number of revolutions of the strip a Veeder counter is mounted on the panel, and is connected directly to the driving shaft of the roller by a flexible lead. In this instrument approximately 9 revolutions of the driving shaft correspond to one revolution of the strip. Also mounted on the bakelite panel in front of the instrument are the counters, the rheostat dials, and the electric switches. Thus all the controls are directly in front of *E*, as he sits with stop watch in hand. He merely gives the starting and stopping commands, times the trial, after adjusting the rheostats for the speed required, and records the reading of the counters. Thus one *E* can easily and accurately control the experiment.

The whole instrument is built on a metal framework and is contained in a wooden box with a removable lid having on top the windows through which *S* sees the tasks. By resting his arm on the top of the instrument *S*, who is standing, is able to work with maximum efficiency. *E* sits in front of the instrument where all the readings are visible. The instrument is portable and is immediately ready for operation by plugging in a cable to a 110 volt main. It is sturdily built and operates without any attention. It is also adapted to a number of tasks other than those described in this article.

Subjects.—The *S*s of this experiment were all male University students ranging in age from 20–26 years. Task I, the complex series, was done by 20 students, 5 of whom repeated the experiment, and 2 others did it three times. Task II, the simple series, was done by 8 *S*s, 4 of whom had taken part in the first task. Task III was done by 16 *S*s, all of whom had undertaken the first task. The data was gathered between the months of October 1932 and March 1933.

All *S*s were aware of the purpose of the experiment and coöperated to the best of their ability. They were stimulated to put forth their maximum effort and they were promised that the records of their scores would be shown them. A good deal of rivalry was manifest as all *S*s lived in the same University dormitory and were well known to each other. To their unflagging efforts to maintain maximum efficiency, no small credit is due, and the author hereby wishes to express his appreciation.

Experimental Procedure.—*S* was shown the instrument, its purpose was briefly explained and a demonstration of Task I was given. Especial care was taken to point out the right from the wrong perforations. Then *S* was told: "Tap only the right dots as they appear in front of the window." The machine was then set in operation at a very slow speed, the switches operating the counters not being turned on. After a brief trial at slow speed, the machine was then set to the correct speed. *S* was told to wait the command 'Ready! Go!' to tap all the right dots as they appeared, and to keep tapping till the command 'Stop!' was given. *E* then switched on the current for the counters, awaited the green flash, cried 'Ready' and, at the red flash, 'Go,' starting the stop watch at the same instant. After 4 complete revolutions the command 'Stop' was given, the watch was stopped and its time in seconds recorded. Also readings of both dials and the Veeder counter were taken and checked. Then the speed was once more adjusted and another trial given. Thus the experiment proceeded till the task was completed. An average time for experimentation in the first task was about an hour. Between each trial *S* was told to relax, and to rest his eyes as much as possible. This rest was necessary as will be seen later.

Preliminary calibration of the instrument and some initial trials (the results of which were discarded) led to a standardised procedure as regards the number of trials

per task and the setting of the rheostat dials to obtain the desired speeds. Starting at slow speeds of 150 seconds for 4 revolutions, 10 trials were given, of increasing speed to the fastest about 45 seconds for 4 revolutions. At this point a rest period of about 5 minutes was given. Then 10 more trials were taken from fastest back to slowest at approximately the same speeds as before. In all about 20 trials of the first task were given to *S* at one sitting. The number of trials varied somewhat, because an attempt was made to get a series of readings throughout the entire speed range. As the speed of the machine fluctuated slightly with the same setting of the rheostat dials, the trials were always timed by a stop watch. Hence some extra trials were given in a few instances.

The two *E*'s, the author and Mr. F. Canty, a graduate student in Psychology, had practiced together till their method of giving the experiment and their recording of the data were as closely as possible alike.

Method of Scoring and of Curve Fitting.—The number of revolutions of the strips varied as follows in this experiment: To all those who had but one session in Task I (Group A) there were given but 2 revolutions of the strip per trial. That is the possible number of correct taps was 50. Those who took two experimental sessions (Group B) had 2 revolutions per trial at their first session and 4 revolutions per trial at their second session. Hence their maximum number of correct taps per trial were respectively 50 and 100. The two candidates who had three experimental sessions on Task I (Group C) had 4 revolutions of the strip per trial. Hence, the maximum number of correct taps for them was 100. In all experimental sessions for Task II (the simple series) the total number of correct taps was 100. In the computations and the tables presented (except Table IX) all trials were reduced to a common basis of 50 correct responses as the maximum number per trial.

For each trial in each session of Tasks I and II, there were tabulated the time per trial, the number of right taps and the number of wrong taps. The score per trial is the difference of rights minus the wrongs expressed as a percentage of the total of rights presented. That is, from the number of right taps recorded on the one dial is deducted the number of wrong taps recorded on the other dial. This difference was changed into a percentage of the total (50 or 100) as the case might be. The percentage scores might accordingly range from 0 to 100 per trial.

From the tabulations of the percentages per trial thus obtained graphs were drawn showing the percentages plotted against time of trial. (Or, as it was labelled, accuracy against speed.) Nearly all curves showed a district tendency to take an ogive form. Accordingly an attempt was made to fit the curves to the standard psychometric function, *i.e.*, the integration of the normal probability curve. This was done in one of the following ways:

(*a*) From the unsmoothed graphs 5 to 7 equidistant points on the time axis were chosen, with their corresponding percentages of accuracy. A psychometric function of best fit was calculated by the method of least squares following the standard procedure and using Urban's weights (**1**, **3**, **16**). This method was used for all sessions of the 5 *S*s who took Task I twice and the 2 *S*s who took it three times, *i.e.* for Groups B and C. Each session yields two curves; one when the speed is changed from slowest to fastest and one when the reverse procedure is followed. Thus were obtained 4 curves from each of 5 *S*s, and 6 curves from each of 2 *S*s.

(*b*) With the object of saving laborious calculations, the abbreviated method of Newhall (**11**, **12**) was followed and his formula (**4**) was used. Accordingly 4 points on the curve were chosen, two below 50 percent accuracy, and two above, from which

the $R_{.50}$ and the h's and σ's were calculated. In a preliminary test with 4 curves, both Newhall's formulæ (3) and (4) were compared in results with those obtained by the method of least squares. The differences were quite small and well within the margin of error. This method was used for the 13 *S*s of Group A who had had but one experimental session with Task I. From these 26 curves were obtained. The advantages of this method were its brevity and the fact that the points chosen did not have to be equidistant on the speed scale.

(*c*) The data obtained from Task II were a trifle too irregular to permit of either of the above methods being used. Hence a graph of percentage of accuracy plotted against times per trial was made. A line of best fit was drawn between points R_{80} and R_{20}; using the values of these points found in this manner Newhall's procedure formula (3) was followed.

Method (*c*) differs from method (*b*) by using as points for calculation of the limen and the precision, points of best fit, rather than actual values found in experiment. As the primary object of the experiment was to determine the types of curve of best fit, rather than the actual constants in any given case, and as the closeness of fit was determined from deviations of the actual values found experimentally from the theoretical curves in accordance with Pearson's χ^2 method, the method of finding the constants is of little significance, but rather one of convenience.

It is well to call attention to several points at this juncture:—All the data obtained in the experimental sessions were used. Naturally there were times when *S* became flustered or distracted and hence his score at that particular speed deviated somewhat from the expected value. The presence of one of these points tends to raise the χ^2 value considerably. Any inversions of the first degree, relatively few in number, were found at these points. These values were not discarded. In these cases two χ^2's were calculated, one omitting the deviating score and one including it. When such procedure was followed, the deviating score was designated by an asterisk. The two sets of values are tabulated to the left and right respectively of Tables I and II.

Despite great care in calibrating the instrument for speed, it was impossible in the experimental sessions to obtain this speed with accuracy. Hence a stop watch was always used for timing. Moreover, there sometimes occurred gaps in the speed scale. In general these were not serious except at mid-points of the scale. Hence arose the difficulty of obtaining equidistant points on the speed scale when applying the method of constant differences to calculate the curves by the method of least squares. Sometimes even with different settings of the rheostat dials, particularly at the faster speeds, identical times were obtained. Sometimes in Task II at critical points identical speeds were deliberately obtained in two or three successive trials. In both instances the percentages for two or three trials at the same speed were averaged. This procedure increased the reliability of the score, but brought in additional practice and fatigue effects. The same criticism, of course, applies to those sessions that had 4 revolutions instead of 2 revolutions of the strip. Greater reliability was achieved at the expense of the added complications due to practise and fatigue.

The initial readings per session had sometimes to be discarded, while *S* was becoming adapted to the experiment. At the fastest speeds and consequently the lowest percentage accuracy, the experimental error became quite marked, and such extreme values were also discarded. In any event, the values at the ends of the distribution are of little importance in calculating the curves. As the range is always given in the tabulations and averaged from 10–90 percent, which is quite satisfactory for a psychometric function, the omission of the extreme values does not materially affect the conclusions.

TABLE I

Closeness of Fit of Curve

Task I (Complex Series)

Groups A, B, and C refer respectively to those subjects who had taken 1, 2, or 3 experimental sessions. Top row for each subject gives the values for the starting curve; the bottom row, for the finishing curve. Range is given in percentiles of accuracy. * indicates percentile value of point omitted in calculation. χ_1^2, $n_1 + 1$ and P_1, — values found for curves with the * points included.

Subjects	Date	Range	χ^2	$n+1$	*	P	χ_1^2	n_1+1	P_1
A									
1	Oct. 18	8–97	11.02	8	82.7	.14	19.25	10	.023
		8–96	4.93	9		.76			
2	Sept. 10	8–89	9.42	10		.40			
		15–96	6.45	9		.60			
3	Sept. 12	17–93	1.63	9		.98			
		14–81	8.85	11		.59			
4	Oct. 21	8–89	10.52	8		.16			
		3–97	15.89	11		.10			
5	Oct. 23	4–89	6.39	7	25.4	.38	18.54	8	.010
		5–96	8.93	9	40.6	.35	26.13	10	.002
6	Oct. 23	9–95	10.47	9		.24			
		9–91	6.20	9		.63			
7	Oct. 30	18–95	6.68	7		.16			
		10–91	2.71	7		.84			
8	Nov. 8	13–84	19.08	11	41.5	.01	32.98	13	.002
		5–90	11.03	10		.27			
9	Nov. 8	6–99	18.72	10	36.2	.03	30.96	11	.002
		7–98	4.47	9		.81			
10	Nov. 10	19–96	9.31	10		.41			
		13–96	14.22	10		.12			
11	Oct. 25	6–93	8.50	9	48.9	.39	25.34	10	.003
		14–93	5.06	9		.75			
12	Jan. 19	16–99	11.79	9		.16			
		15–99	8.70	6	26.3	.12	20.53	7	.002
13	Jan. 20	11–94	7.56	7		.27			
		18–97	5.27	7		.51			

TABLE I—(*Continued*)

Subjects	Date	Range	χ^2	$n+1$	*	P	χ_1^2	n_1+1	P_1
B									
14	Oct. 23	25–92	4.64	9	83.5	.79	14.580	10	.10
		10–96	6.39	9		.60			
	Jan. 25	4–99	4.94	7		.55			
		6–92	7.09	6	54.0	.22	11.930	7	.06
15	Nov. 26	19–99	3.12	8		.85			
		15–78	4.86	6		.44			
	Jan. 26	5–99	2.17	7		.90			
		12–89	4.20	6	56.0	.52	12.236	7	.06
16	Nov. 1	10–88	8.60	9		.38			
		6–100	3.64	9		.86			
	Jan. 27	20–95	3.88	8		.79			
		19–66	9.58	8		.22			
17	Nov. 3	16–74	2.22	7		.90			
		16–92	2.97	8		.88			
	Jan. 23	11–91	5.91	6		.41			
		5–96	4.76	8		.69			
18	Oct. 17	19–86	11.00	8	57.5	.14	19.600	9	.01
		14–83	6.06	11	35.0	.81	12.420	12	.33
	Jan. 16	9–83	5.41	8		.61			
		11–88	6.36	8		.50			
C									
19	Jan. 20	11–93	6.98	5	47.6	.14	18.52	6	.002
		14–97	2.54	7		.86			
	Feb. 2	10–72	9.96	9	28.5	.27	19.88	10	.030
		15–85	6.48	11		.77			
	Feb. 12	19–95	10.83	11	53.0	.37	17.68	12	.090
		21–88	6.35	11		.79			
20	Jan. 18	20–77	2.95	6		.71			
		23–89	3.86	6		.57			
	Feb. 1	28–82	12.23	10		.20			
		15–79	5.01	11		.89			
	Feb. 20	20–78	8.69	9		.37			
		20–80	14.42	10		.11			

TABLE II

CLOSENESS OF FIT OF CURVES

Task II (Simple Series)

Range is given in percentiles of accuracy. * indicates percentile value of point omitted in calculation. χ_1^2, $n_1 + 1$ and P-values found for curves with the * points included.

Subjects	Date	Range	χ^2	$n+1$	*	P	χ_1^2	n_1	P_1
14	Mar. 16	13–90	8.44	8	68	.30	16.70	9	.04
16	Mar. 9	13–69	10.75	11		.38			
19	Mar. 16	14–52	13.46	14		.41			
20	Mar. 8	3–79	15.14	14	62	.26	25.00	15	.04
21	Feb. 2	24–96	11.88	10	62	.22	16.18	11	.10
22	Feb. 7	22–99	7.16	9		.52			
23	Feb. 9	13–81	8.16	7		.22			
24	Feb. 16	16–84	10.33	10	70	.33	20.98	11	.02

RESULTS

When inspection of the preliminary graphs of Task I indicated the possibility of the relationship of speed and accuracy fitting a psychometric function it was planned to calculate the corresponding theoretical curves, and to utilize Pearson's χ^2 criterion to determine the closeness of fit. This has been done; in Tables I and II are given the χ^2's and P's for 58 curves of Task I and 8 curves of Task II. The constants of the curves are given in Tables III and IV. The method of selecting the curves of best fit has been described in a previous section. The testing of the fitting of the curve was done by means of the formula

$$\chi^2 = kS\frac{(p^1 - p)^2}{p'q'}$$

as described by Brown and Thomson (**3**, pp. 79–81 and **6**, p. 71).

The values of P corresponding to χ^2 were taken from Pearson's Tables (**13**, pp. 26–27) using the entries in the $n + 1$ row, where n is the number of points fitted on the curve. Tables I and II also give the range of the curves from which the χ^2's were calculated and the number of points used in drawing the curves. As has been previously pointed out all the values found in the experimental sessions were used, with

the exception of values at the extremes of the accuracy scale, where adaptation to the experiment, fatigue, or undue margin of error hold too much weight to make the values trustworthy. Within the ranges indicated all points were used except for

TABLE III

COMPARISON OF STARTING AND FINISHING CURVES. TASK II

Groups A, B, and C indicate those who had taken respectively 1, 2 or 3 experimental sessions on Task II. The numbers (1), (2) and (3) indicate the 1st, 2nd or 3rd session. The units are seconds. $R_{.50}$, $R_{.1487}$ and $R_{.8513}$ refer to the accuracy percentile.

Subject	Starting Curve					Finishing Curve				
	$R_{.50}$	$R_{.1487}$	$R_{.8513}$	h	σ	$R_{.50}$	$R_{.15}$	$R_{.87}$	h	σ
A(1)										
1	50.41	36.85	63.90	0.0524	13.5	49.07	36.83	61.3	0.0579	12.2
2	43.34	34.78	51.90	0.0822	8.6	42.40	34.50	50.3	0.0875	7.9
3	42.41	33.73	51.10	0.0812	8.7	39.17	31.11	47.2	0.0882	8.0
4	57.20	38.34	76.20	0.0374	18.9	51.16	40.32	62.0	0.0655	10.8
5	52.21	39.74	63.70	0.0589	12.0	50.95	39.36	62.5	0.0609	11.6
6	51.50	37.59	65.30	0.0508	13.9	47.88	35.17	60.6	0.0557	12.7
7	47.15	35.25	59.10	0.0594	11.9	43.22	35.60	51.8	0.0872	8.1
8	60.22	36.17	84.40	0.0293	24.1	48.98	39.20	58.8	0.0721	9.8
9	54.62	44.39	64.80	0.0693	10.2	46.61	37.55	55.7	0.0776	9.1
10	50.60	33.65	67.60	0.0416	17.0	47.80	34.43	61.2	0.0527	13.4
11	50.40	37.70	63.10	0.0566	12.7	48.34	33.39	63.4	0.0471	15.0
12	49.62	34.48	64.90	0.0465	15.1	41.94	33.25	50.7	0.0812	8.7
13	46.55	34.75	58.30	0.0598	11.8	45.00	31.50	58.3	0.0527	13.4
B(1)										
14	45.23	30.03	60.43	0.0465	15.2	44.88	34.88	54.88	0.0710	10.0
15	48.00	33.20	62.80	0.0480	14.7	41.20	32.50	49.90	0.0810	8.7
16	56.00	43.40	68.60	0.0439	6.2	49.94	39.04	60.84	0.0652	17.7
17	43.52	31.86	55.22	0.0604	11.7	43.36	31.36	55.36	0.0592	11.9
18	54.20	39.80	69.00	0.0478	15.5	47.60	33.60	61.60	0.0506	14.0
C(1)										
19	44.44	36.24	52.64	0.0862	8.2	42.73	33.33	52.13	0.0754	9.4
20	43.55	31.15	55.95	0.0572	12.3	40.71	30.21	51.21	0.0676	11.4
B(2)										
14	41.14	32.94	49.34	0.0860	8.2	39.11	30.01	48.21	0.0784	9.0
15	38.25	30.05	46.45	0.0862	8.2	37.50	27.50	47.50	0.0706	10.0
16	40.50	28.70	52.30	0.0602	11.7	34.25	20.65	47.85	0.0520	13.6
17	38.40	28.50	48.30	0.0716	9.9	35.88	26.68	45.08	0.0774	9.1
18	39.30	28.00	50.60	0.0622	11.3	38.12	26.42	49.82	0.0604	11.9
C(2)										
19	40.84	32.24	49.44	0.0824	8.6	39.35	30.05	48.65	0.0760	9.3
20	34.43	18.13	50.73	0.0434	16.3	33.34	20.44	46.24	0.0548	12.9
C(3)										
19	35.25	25.15	45.35	0.0704	10.0	34.63	26.93	42.33	0.0920	7.7
20	32.95	19.45	46.45	0.0526	13.4	32.02	22.32	41.72	0.0728	9.7

TABLE IV

CONSTANTS OF CURVES OF BEST FIT

Task II (Simple Series)

$R_{.50}$ equals accuracy value of 50 percent. h equals coefficient of precision. σ equals standard deviation.

Subjects	$R_{.50}$	h	σ
14	21.24	0.1860	3.8
16	25.00	0.1170	6.0
19	20.30	0.1230	5.8
20	22.00	0.1414	5.0
21	23.50	0.1414	5.0
22	22.50	0.1570	4.5
23	24.84	0.2830	2.5
24	24.50	0.1080	6.5

those frequencies which are listed in the column headed by an asterisk; this omission of one point only occurs in 15 curves of the 58 of Task I and in 4 curves on 8 of Task II. In these cases the χ^2's and P's are given. It is readily seen that the one point makes a disproportionate difference in the closeness of fit. These discrepancies are probably due to distractions, change of attitude or method on the part of *S*, or to emotional disturbances particularly confusion caused by making of errors.

Considering the pressure under which *S*s were working the agreement of the theoretical with the observed values in the curves is quite remarkable. All the curves but one would fall in the category designated by Fisher (**6**, p. 82) 'as giving no reason to suspect the hypothesis tested.' The distribution of the values of P indicating the closeness of fit of the curves is given in Table VI. The mean value of the 66 P's in both Tasks is about 0.47 which is quite satisfactory.

In order to test whether the general level of variability of the χ^2 values conforms with expectation, the criterion described by Fisher (**6**, p. 61) was made use of. Fisher states: "The sum of the independent values of χ^2 is itself distributed in the manner shown in the table of χ^2, provided we take for n the number $S(n)$ calculated by adding the several values of n for the separate experiments. . . . A good approximation (of the testing of this new value of χ^2) is given by assuming that

$(\sqrt{2\chi^2} - \sqrt{2n-1})$ is normally distributed about zero with unit standard deviation. If this quantity is materially greater than 2, the value of χ^2 is not in accordance with expectation."

As has been already stated 15 of the 58 curves of Task I, and 4 of the 8 curves of Task II had one point omitted in the curve fitting. The χ^2 values given to the left of Tables I and II are on this basis. When the points are included the χ^2 values are given to the right of Tables I and II. Using the values of χ^2 and n to the left of Tables I and II (*i.e.* omitting the percentages in the asterisk column), the criterion $(\sqrt{2\chi^2} - \sqrt{2n-1})$ works out

for Task I	−1.9
for Task II	0.5
for both tasks	−1.7,

for the values of χ^2 and n to the right of both tables, the value of $(\sqrt{2\chi^2} - \sqrt{2n-1})$ is

for Task I	2.37
for Task II	2.24
for both tasks	3.16

For the curves as originally tabulated, *i.e.* with the disputed point omitted, the criterion is completely satisfied. When all points are included the values of $(\sqrt{2\chi^2} - \sqrt{2n-1})$ are but slightly in excess of 2.

The evidence therefore seems quite ample that the relationship between speed and accuracy in the motor tasks of this experiment may be expressed by a psychometric curve, that is the integration of the normal probability curve. Naturally the constants of the curve vary with the individual *S*, the variability of whom is still further influenced by practice and fatigue, and similar factors. This point will be discussed later.

In Table V are presented the results of Task III. In this task which was given last, *S*s were made to tap at maximum speed for 30 seconds, followed by a 30 second rest, during 10 minutes. The readings were taken at every half minute period and checked at the completion of the session. The averages for the 10 trials are given in the table. The time per

trial, 30 seconds, was chosen as Ss could maintain their maximum pace without showing fatigue during this interval. It agreed also with the mean time per trial of the other tasks.

TABLE V

RESULTS OF TAPPING TEST

The subjects tapped for 10 minutes, alternately working and resting for ½ minute intervals.

Subjects	Date	Mean	M.D.	σ
1	Mar. 26	83.5	2.4	2.9
5	Mar. 26	91.9	7.9	10.1
6	Mar. 26	96.6	4.1	5.4
7	Feb. 16	102.2	6.0	7.2
8	Feb. 16	107.6	5.0	5.8
9	Mar. 26	90.6	3.4	3.9
10	Mar. 1	95.8	4.8	8.1
11	Feb. 27	103.4	6.7	8.2
12	Mar. 5	99.1	14.7	22.1
14	Feb. 27	117.4	6.2	7.2
15	Mar. 24	88.1	6.3	10.3
16	Mar. 24	102.7	5.9	7.8
17	Mar. 5	101.1	4.7	5.9
18	Mar. 1	102.0	7.2	9.6
19	Mar. 5	114.7	9.5	11.6
20	Mar. 1	120.6	6.1	7.1

While fatigue did not bother Ss there was some tendency to blocking. Moreover the stylus tended to slip into the wells when the localisation was slightly inaccurate. However as the bottom of the well was slightly lower than the brass plate, the gain in accuracy was probably compensated for by the loss in speed. It would have been much better to have the stationary tapping done on a perforated oilcloth strip as in the other tasks.

Practice and Fatigue.—By practice is here meant any combination of factors that tends to improve the performance while the external working conditions remain constant. It accordingly includes adaptation to the task, warming-up effect, subduing of emotional tension, etc., besides the amelioration due to mere repetition of the task. On the other hand, by fatigue is meant any combination of factors that tends to decrease the efficiency of the performance, while the external working conditions remain constant. It, of course,

TABLE VI

FREQUENCY DISTRIBUTION OF VALUES OF *P*, INDICATING CLOSENESS OF FIT OF CURVES

Range	Task I	Task II	Total
90–100	3		3
80–99	7		7
70–79	7		7
60–69	5		5
50–59	6	1	7
40–49	4	1	5
30–39	7	3	10
20–29	7	3	10
10–19	10		10
0– 9	2		2
Totals	58	8	66
Means	48.6	35.0	47.0
σ Dist	37.0	10.0	26.0

embraces in this wider meaning of the term conditions such as decreasing physical efficiency, lassitude, boredom, tension, and muscular cramping. Both practice and fatigue are operating simultaneously; though practice is more readily perceived at the outset of the task and fatigue towards its termination. Comments of *S*s give clues as to the strength of the two factors. Certain semi-objective signs of fatigue are particularily noticeable such as hand trembling, cramped muscles and most of all, eye strain and after images of motion. An attempt is made to compensate for fatigue effects by telling the *S* to relax as much as possible between trials. Eye strain and after images of motion are partially avoided by requesting *S* not to look at the moving strip until just before starting each trial of the tasks. In some cases the experiment had to be stopped for a few minutes while *S* rested. At the middle of each session, while changing from increasing to decreasing speed, there was always given a five minute rest period. In spite of these precautions, real fatigue of *S*s was quite evident, and some complained of after images of motion. It is remarkable that *S*s were able to maintain such a satisfactory level of efficiency, as evidenced by the regularity of their performance curves. The discard-

ing of one point in some of the curves seems quite a legitimate procedure in view of the strain placed upon Ss by the tasks. Moreover, the first few trials, which were always taken at very slow speeds, were used to adapt S to the experiment. In any event they were outside the critical range for the determination of the performance curve.

Besides these semi-objective evidences of practice and fatigue, the data enable one to form a fairly accurate idea of their effect. One way of showing the practice effect is by a comparison of the starting and finishing curves of each session. Three points on each curve have been singled out. These are the points corresponding to $R = .1587$ and $R = .8413$ and $R = .5000$. They have been found by Newhall's method (**12**), when determining the sigmas, and they are well within the effective range of the data. At the outset of each session while the speed is increasing the point $R = .8413$ would come first followed by $R = .50$, then $R = .1587$. With decreasing speed the reverse order holds. Consequently the greatest practice effect in any session should be shown between the $R = .8413$ points and least between points $R = .1587$.

The average values of these points for the beginning and finishing curves of the 20 Ss of groups A, B, and C, are given in Table VII. They give the information as to the practice effect in the first session of Task I. As one would expect practice is most evident between the points corresponding to $R = .8413$. In other terms the efficiency required to

TABLE VII

PRACTICE EFFECT OF ONE SESSION

Differences Between Identical Points on the Starting and Finishing Curves

	$R_{.1587}$	$R_{.5000}$	$R_{.8413}$	h
Starting	36.16	49.52	62.95	0.0558
Finishing	34.86	45.65	56.48	0.0678
Difference	1.30	3.87	6.47	−0.0120
Percent gained	3.67	7.80	10.30	21.6000

The units are given in seconds for the time of 2 revolutions of the strips, (fifty rights presented). Number of subjects = 20.

produce an accuracy of 84.13 percent is achieved in the finishing curve at a speed 6.47 seconds faster for 2 revolutions than in the starting curve. This corresponds to a 10.3 percent increase in speed at the same efficiency. The corresponding percentage speed increases for points $R = .5000$ and $R = .1587$ are 7.8 percent and 3.6 percent. At the same time there is an increase of the coefficient of precision (or a reduction in variability) by about 21.6 percent. Similar effects have been found by other investigators in data to which the constant process method has been applied (**3**, **4**, **5**, **7**).

When we compare in Table VIII the averages of the 2 Ss who had 3 sessions, from each of whom 6 curves were obtained, in general the same effect of practice is seen. Keeping in mind that $R = .8413$ is the first of the 3 chosen points to be presented in the starting curve and the last in the finishing curves, Table VIII shows a practice effect, similar to that

TABLE VIII

AVERAGE PRACTICE EFFECT OF THREE SESSIONS

S = Starting Curve. F = Finishing Curve. R_{1587}, R_{50} and R_{8413} points on curves corresponding to accuracies of 15.87 percent, 50 percent and 84.13 percent. D = Differences between successive curves. h = Coefficient of precision. Number of subjects = 2. Units are given in seconds for the time of 2 revolutions of the strip (fifty rights presented).

Curve	$R_{.1587}$	Diff.	$R_{.50}$	Diff.	$R_{.8413}$	Diff.	h	Diff.
S_1	33.70		44.00		54.30		0.0717	
F_1	31.77	1.93	41.72	2.28	51.67	2.63	0.0715	0.0002
S_2	25.19	6.58	37.64	4.08	50.09	1.58	0.0629	0.0086
F_2	25.25	−0.06	36.35	1.29	47.45	2.64	0.0654	−0.0025
S_3	22.30	2.95	34.10	2.25	45.90	1.55	0.0615	0.0039
F_3	24.63	−2.33	33.33	.77	42.03	3.87	0.0824	−0.0209

seen in Table VII. Thus in the difference columns starting with row F_1, are shown alternate increases and decreases in speed gain as one proceeds from $R = .1587$ to $R = .8413$, according to the amount of practice.

The progressive increase in h, the coefficient of precision, is not so evident in this table, probably for the reason that while practice tends to increase the precision, the increasing speed of the latter series tends to lower it.

At two points in the table there is a decrease in speed. (The differences are negative.) They occur at the outset of finishing curves (2) and (3). It is to be remembered that a rest period occurs between the starting and finishing series of each session: Here *S* would be more likely to show a warming up effect, just after a rest from a fatiguing task, and this is what probably occurs.[1]

The curves of the two *S*s, numbers 19 and 20, who had taken 3 experimental sessions, are given in plates I and II.

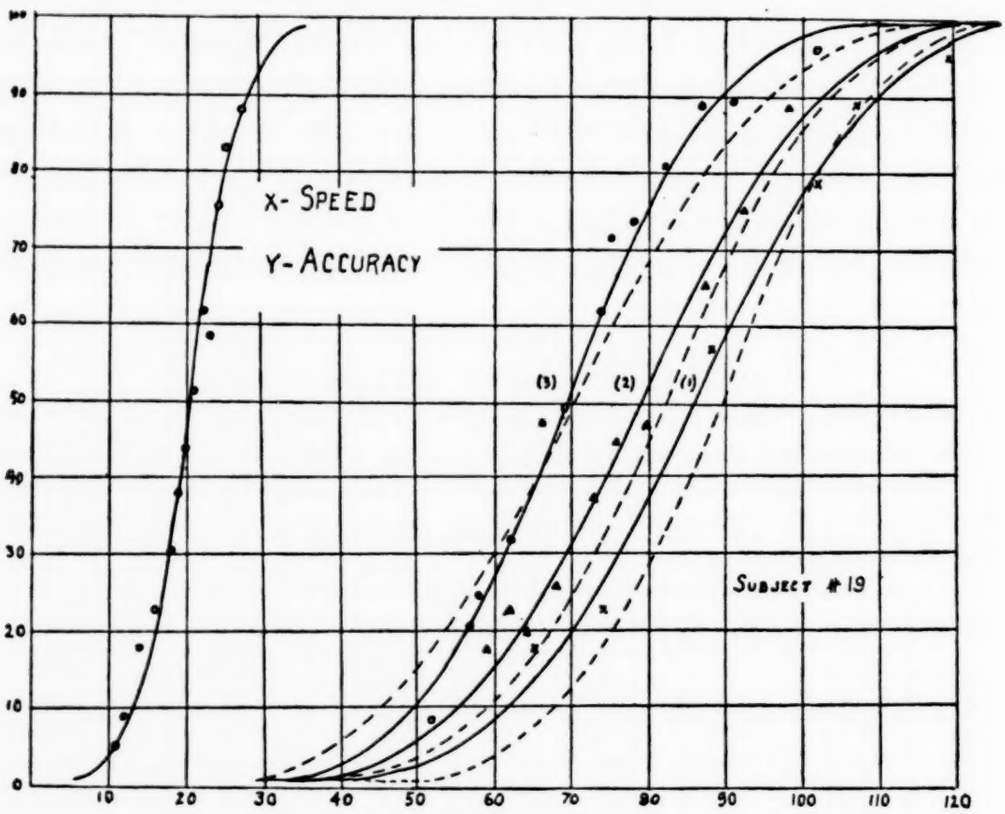

PLATE I

The points corresponding to the curves are shown only for the 3 finishing curves for the sake of clearness. There is also indicated on the plate the curve of each *S* for Task II. The plates show quite well the general form of the curves, the effect of practice, and the conformity of the observed to the theoretical points. The results for these two *S*s are typical of the general findings.

[1] Data about to be published by the author on the effects of continuous work bring out this phenomenon quite conclusively.

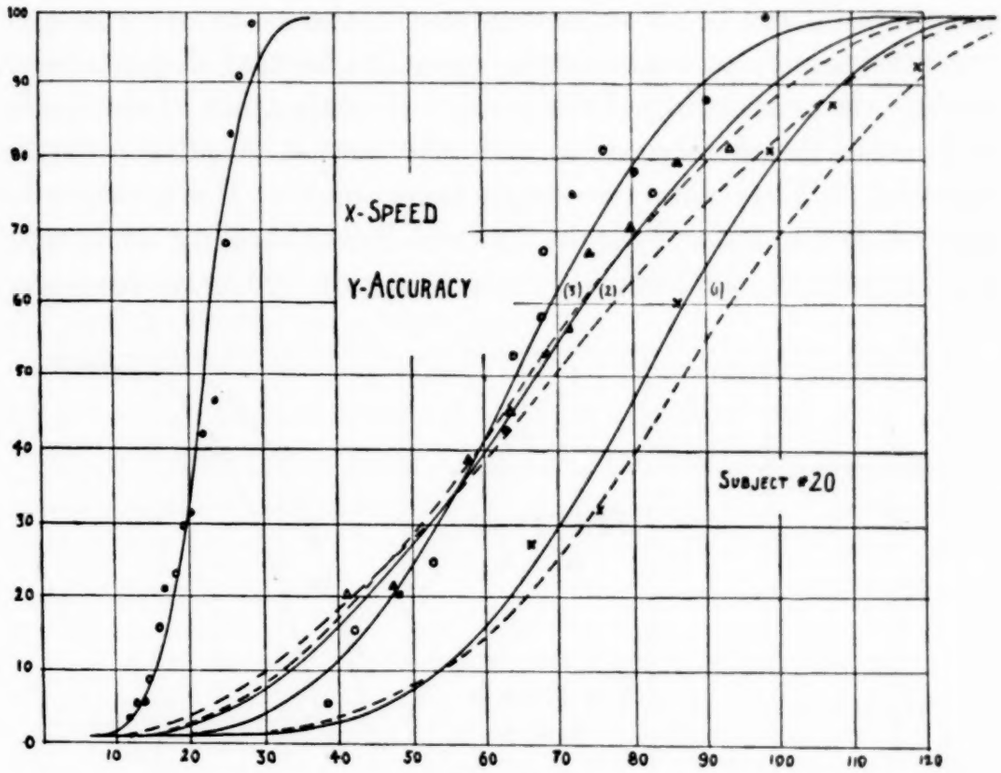

PLATE II

Plate III gives the starting and finishing curves of a typical subject who had but one experimental session in Task I. Three points at the beginning of the starting curve and one point on each curve at the fastest speed are not given, as they are on the tail of the distribution.

This curve shows how the practice effect may be obtained graphically. First, a general measure is given by the abscissae difference corresponding to an accuracy of $P_{.50}$. Secondly, the practice effect at any speed is given by the ordinate difference at that speed; it is progressively larger as the speed becomes slower. These two measures of practice correspond to the coefficients of coercion and retentivity in the hysteresis curve of magnetism to which the speed-accuracy curve resembles in structure and, to some extent, in function.

Analysis of the Factors Involved in the Task.—A preliminary attempt was made on the basis of the experimental data, to determine the component factors of the tasks. It is readily seen that the psychological factors become more intricate as we proceed from Task III to Task I. The simple

tapping of Task III is relatively static, the areas to be tapped are stationary, and involve a constant lateral displacement of the arm or swaying of the body. Localisation of the wells, and rapidity of movement are the major requirements of this task. The other two tasks are dynamic; the areas to be tapped are constantly moving and are changing in lateral displacement. In Task II, the perception of the perforations

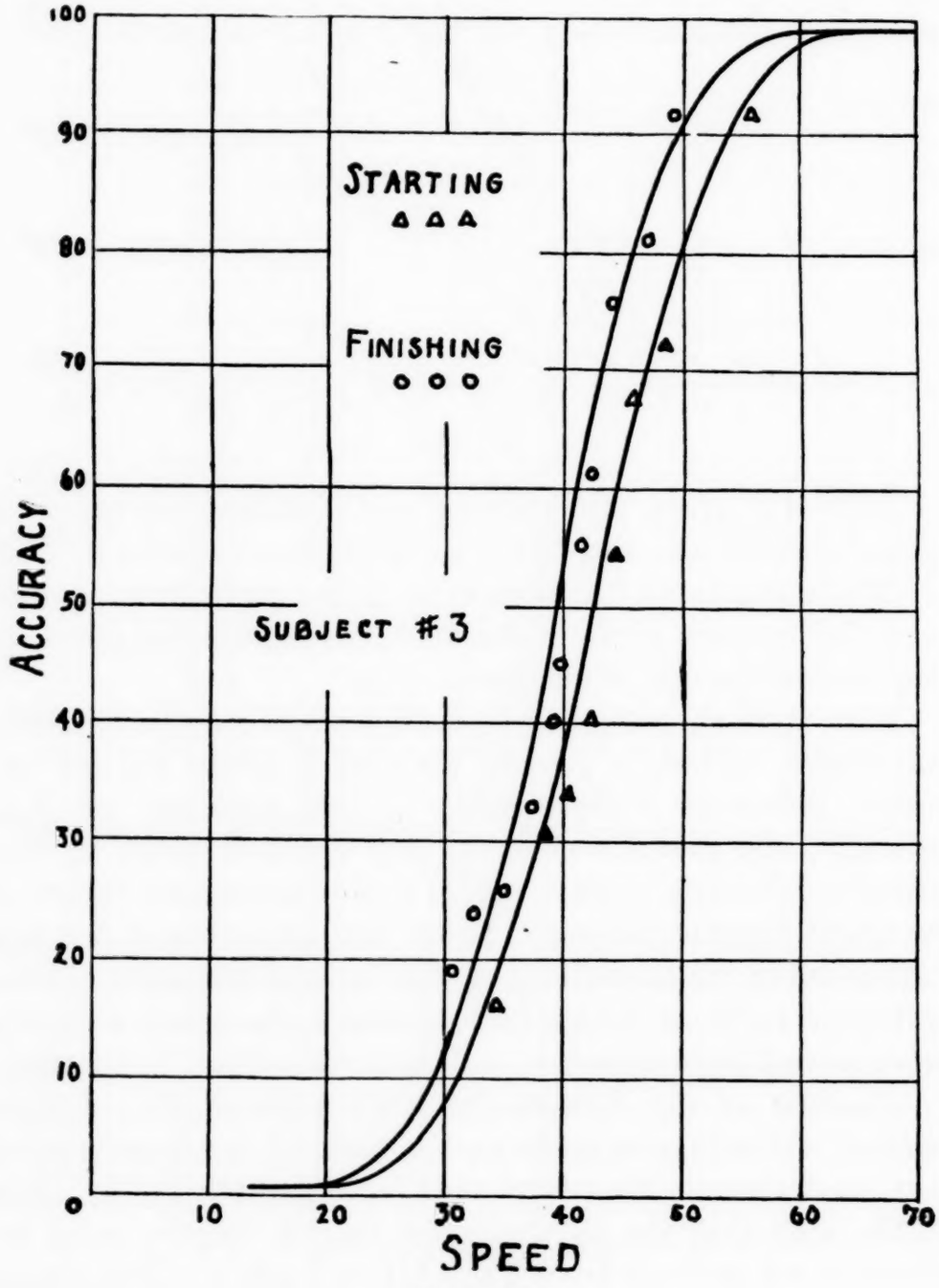

PLATE III

and the neuro-muscular adjustments required properly to time and to control the movements are additional factors which make this task more complicated than simple tapping. Finally in Task I, discrimination of the rights and the wrongs is required. In Task II the perforations were of a larger area (1.833 times that of Tasks I and III), a narrower average displacement (1″ instead of 2.75″) and of a different shape, rhomboid instead of rectangular.

Hence it is to be expected that we would find progressively longer time for the same number of possible contacts, as we go from Tasks III to Task I. Such is indeed the case, and the times required for the tasks give us a clue to the complexity of the operation. Table IX, in which all tasks are reduced to

TABLE IX

COMPARISON OF THE ABSOLUTE SPEEDS OF THE THREE TASKS

I, II and III give the times in seconds of the $P_{.50}$ accuracy for Tasks I, II and III estimated on the basis of 150 possible contacts. I is based on the average of the starting and finishing curves of the first session of Task I. A, B, C, give the percentage proportion respectively of II/I, III/I, and III/II. P_1, P_2 and P_3 give the accuracy percentile values of Tasks I and II to which the times of III correspond and for Task I to which the times of II correspond.

Subject	I	II	III	A	B	C	P_1	P_2	P_3
14	90.11	63.72	38.3	71	43	60	2.8	1.9	14.2
16	106.00	75.00	43.8	71	41	59	1.3	4.3	12.5
19	87.16	60.30	39.3	69	45	65	1.0	10.5	6.6
20	84.26	66.00	37.3	78	44	57	2.0	2.8	21.3
1	99.48		54.0		54		3.7		
5	103.16		49.0		47		1.0		
6	99.38		46.6		47		2.3		
7	90.37		44.0		49		1.4		
8	109.20		41.8		38		2.6		
9	101.23		49.7		49		0.4		
10	98.40		47.0		48		4.5		
11	98.74		43.5		44		2.4		
12	91.56		45.4		50		2.7		
15	89.20		51.1		57		5.0		
17	86.88		44.4		51		3.7		
18	101.80		44.2		44		2.6		
Averages	96.00	66.26	44.96	72.00	47.00	60.00	2.46	4.90	13.70
σ Dist.	7.30	5.44	4.44	3.42	4.65	2.95	1.23	3.36	5.23
$\frac{100\sigma}{M}$	7.64	8.22	9.88	4.74	9.91	4.90	50.00	70.60	38.20

basis of 150 possible contacts, gives the 50th percentile accuracy of Tasks I and II, and the average time for stationary tapping of Task III. From columns B and C, we find that on the average Task III takes but 47 percent of the time required to do Task I, and 60 percent of the time of Task II; Task II required 72 percent of the time required to do Task III. These percentages are fairly constant among the 16 *Ss* (only 4 of whom however did Task II). Of course the values found are but tentative, owing to the small number of *Ss*, and the variation of areas, shape and displacement in Task II. One may assert, however, that the time to achieve a constant accuracy, say 50 percent, varies directly with the complexity of the task.

It is interesting to determine what accuracy would be given in Task I by the speeds listed in II and III. We find that the average accuracy yielded by the tapping time of Task III, gives an accuracy as low as 2.5 percent in Task I, and 4.9 percent in Task II. It is to be noted that the probable accuracy of random tapping at the speed of Task III, under the assumption that the probability is proportional to the respective areas of the perforation and of the oilcloth strip exposed, would yield an accuracy of about 1.4 percent in Task I, and 7.3 percent in Task II. These values according check quite well, and the agreement between *Ss* is also surprisingly good. The assumption of course does not take into account the tendency to tap the central perforations more frequently nor does it allow for the possibility (shown by some *Ss*) of tapping both right and wrong perforations at high speeds.

The correlation coefficient between the speeds of the P_{50} accuracy values of Tasks I and the means of Task II was determined for 16 *Ss*. Using Spearman's rank difference method the ρ was found to be 0.26 with a probable error of .16. The value found, though unreliable, is but slightly lower than that usually found in motor abilities (**9**). The caution given by Kelley (**8**, p. 221) that the coefficient of correlation should be interpreted in the light of the range of the traits measured is here applicable.

Relationship Between Accuracy and Errors.—Accuracy has been measured in this experiment by the difference between the correct taps and the errors made at each speed. There are several possible kinds of inaccuracies: (*a*) Correct perforations which are not tapped, *i.e.* the omissions, and (*b*) the wrong perforations which are tapped. Only the second type are designated as errors in the tabulations. A third possible type of mistake is not recorded, those which consist of an unsuccessful attempt to tap the wrong type of perforations. In this case there is a double failure—one of discrimination and one of coördination. In Task I, the number of wrong perforations is just one half the number of correct perforations. Hence the probability of an error being of the second or third type is not greater than one on three.

One would expect relatively few errors to be made at the slowest speeds and greatest accuracy. Moreover, at the very fastest speeds the number of errors is greatly reduced due to difficulties of coördination—few of both right and wrong perforations are tapped. The most errors accordingly occur at medium accuracies. Hence the correlation between accuracy and error would not be expected to be a linear one. And in general this is what we get. Table X gives the correlations found (both Pearson Product Moment r's and η's)

TABLE X

CORRELATION BETWEEN ERRORS (x) AND ACCURACY (y)

N —Number of Experimental Trials.
M_x and M_y—Mean values for x and y variables.
σ_x and σ_y —Standard variation of Distributions.
r_{xy} —Pearson Product Movement Correlation and Its Probable Error.
η_{xy}, η_{yx} —Curvilinear Correlation of Variables.
B —Blakeman's Criterion for Linearity.

	N	M_x	M_y	σ_x	σ_y	r_{xy}	η_{yx}	B	η_{xy}	B
IS	253	2.78	54.65	2.93	34.2	−0.360±0.037	0.358±0.039	L	0.441±0.033	C
IF	228	2.25	59.65	2.81	33.6	−0.455±0.035	0.524±0.032	C	0.524±0.032	C
No. 19	62	7.27	61.24	5.41	26.7	−0.519±0.062	0.701±0.044	C	0.727±0.039	C
No. 20	67	8.81	62.91	4.13	25.1	−0.032±0.081	0.283±0.076	L	0.510±0.061	C

IS and IF—Starting and Finishing sections of Task I.
First Session—20 subjects.
No. 19 and No. 20—All Data, 3 sessions, for subjects 19 and 20.

between Accuracy and Errors. All the trials from the first experimental session of Task I for the twenty Ss were used as data for the starting and finishing sections of each session. Correlations were also obtained from the data of the two Ss (numbers 19 and 20) who had taken three experimental sessions in Task I. It is to be seen that in general the correlations are curvilinear, negative and significant. There is a fair inverse correlation between accuracy and errors, in the sense defined in this paper for these terms.

Discussion

(*a*) *The Application of the Phi-Gamma Hypothesis to the Speed-Accuracy Relationship.*—Psychometric methods have been applied to many problems outside the traditional field of sensory measurement (**14**). Though the integral of the normal probability curve has frequently been fitted to the data of these problems, as far as the author is aware this is the first occasion on which it has been used to express the speed-accuracy relationship in a motor task. Yet fundamentally the same conditions are operating in this case as hold with the discrimination of weights for example. Thus in the lifted weight experiment the limen for 'heavier than' is determined by the value of the weight which yields 50 percent correct judgments when compared with the standard. Weights which are heavier than this yield progressively higher percentages of correct judgments, and weights lighter than the threshold weight give fewer and fewer correct judgments, according to their difference from the 50 percent weight. When a curve showing the percentage of correct judgments is plotted against the weight stimuli it conforms to the Phi-Gamma hypothesis, that is, the curve of best fit is an integration of the normal curve (**2**).

A similar situation is involved in the relationship between speed and accuracy. There is a speed for each *S* which yields a 50 percent accuracy of correct contacts. Speeds greater than this yield progressively smaller percentages of accuracy and slower speeds give increasing percentages of accuracy. In both experiments there is a threshold value of

the stimuli yielding 50 percent of correct scores—and equal deviations from this threshold value give approximately symmetrical variations of percentage accuracy on either side.

Thomson (**15**) has deduced the formulae of the Constant Process Method directly from the assumptions that, (*a*) for every stimulus (generally lifted weights, in this case, speed) there is a constant probability that the reaction in question (correct judgment in general, or, in this case, accurate tapping) will occur; (*b*) that these probabilities are functions of the stimulus given by the integral of the normal probability curve; and (*c*) that the observed values differ so little from the true probabilities that certain mathematical approximations may be made.

Characteristic of the Constant Process Method is the giving of a considerable number of repetitions of each value of the stimulus, generally 100 judgments of a weight, the equal spacing of the different values of the variable stimulus, and the elimination of extreme stimuli thus avoiding the 'tail difficulty.' Where the distribution is not significantly skew, and the data are not heterogeneous the Constant Process Method is probably the best for symmetrical distributions (**3**, p. 75). In the application to this experiment the 100 repetitions of each value of the stimuli (tapping of 100 perforations at each speed) were obtained quickly and easily. Slight variations of the speed at each setting of the rheostats prevented an accurately symmetrical spacing of the speed stimuli. Allowance however was made for this by the methods described on pp. 45 and 46. Extreme stimuli were not included in the calculations due to the large experimental error in determining them.

The chief difficulty in applying the Constant Process Method to this experiment was in the control of factors such as practice, fatigue and emotional blocking. Where emotional effects occurred, there was generally a considerable divergence of the observed from the theoretical values and sometimes inversions of the first degree occur. These large deviations were not numerous and are indicated in the tables. The

practice effect may be partly controlled in the lifted weight experiment by presenting the weights in random order. In this experiment however, for each speed setting of the instrument, the 100 repetitions of the stimulus (dotting of the perforations) are done in succession. Naturally a practice effect follows which becomes more and more evident as we proceed from the first to the last speed settings. As a consequence there is a constant departure of the curve from the integral of the normal Gaussian curve, giving the integral of some negatively skewed curve. Hence Pearson's type III curve when integrated, or the Gram Charlier series might yield a better fit (**10**). The fit of the normal probability curve, however, is sufficiently good to yield a first approximation to the data. Naturally the threshold of discrimination and the coefficient of precision (which is a measure of variability) will vary with each *S*. The *S*s of this experiment were so homogeneous as to age, ability, etc., that the values obtained for each *S* differ but little from the group average as may be seen from Table IX. In general, Pearson's Coefficient of Variation ($100\sigma/M$), is relatively low.

(*b*) *Effect of Complexity of the Task and the Form of the Curve.*—Only three tasks were investigated in this experiment. According to degree of complexity they may be rated in decreasing order, Tasks I, II and III. A study of the typical curves from each group suggests the following tentative hypothesis: The speed-accuracy relationship may be expressed by an equation the curve of which may be fitted to an integral of the normal probability curve. The more complex the task, the slower the speed corresponding to a 50 percent accuracy and the smaller the corresponding h or coefficient of precision. That is, complex tasks give curves which are relatively flat having a small slope and hence showing a wide range of speeds between 0 and 100 percent accuracy. The simpler the task, the greater the slope of the curve and the smaller the speed range between 0 and 100 percent accuracy. Extremely simple tasks, such as continuous tapping of a stationary target, follow practically an all or none law, the accuracy is either 100 percent or 0. We may take this to be a

limiting case, and state that within this range, the slope tends to become infinite and the variability tends to become zero.

The hypothesis tentatively advanced here would lead one to infer that with very difficult tasks speed tests and power tests are identical, for the relationship between speed and accuracy is approximately linear over a wide speed range in tasks of this type. Fairly simple tests show no such identity between speed and power, because for them there is such a small range for which the speed-accuracy relationship is approximately linear.

(*c*) *Best Measure of Speed.*—The non-linear relationship between speed and accuracy found in the experiment brings up the question of what constitutes the best measure of speed when accuracy varies. It is evident that in such tasks any measure of speed dependent upon perfect accuracy involves the selection of a point on the tail of the corresponding frequency function. The same objection applies to a measure of speed which yields 0 percent accuracy. Seemingly some such point as corresponds to the limen of discrimination, that is to the 50th percentile of accuracy, would yield a more reliable measure of speed, one less subject to experimental error, and capable of being calculated with some precision. It goes without saying that some measure of dispersion, either σ or h, the coefficients of variability and precision, should likewise be given. Accordingly it is recommended that the speed of any individual in a motor task involving possibility of error should be characterised by his speed at 50 percent accuracy and his variability (h or σ) in the task. It is felt that much of the confusion of results of speed investigations, particularly the lack of agreement as to the intercorrelations of motor speeds, will be cleared up by the adoption of this definition.

Summary

1. A self-registering apparatus is described to study speed and accuracy in three motor tasks of increasing complexity.

2. The relationship between speed and accuracy in these tasks conforms to the Phi-Gamma hypothesis. The curve of best fit is an integration of the normal probability curve.

3. Constants are given for 66 curves obtained in this experiment.

4. The closeness of fit of the observed values to the theoretical ones has been checked by Pearson's χ^2 method, and by his criterion $\sqrt{2\chi^2} - \sqrt{2n-1} < 2$.

5. Practice and fatigue effects are estimated.

6. A preliminary analysis of the component factors of these tasks is given, and a comparison is made of the absolute speeds of the tasks.

7. The correlation between accuracy and errors is curvilinear, and there is an inverse relationship between them.

8. The complexity of the task determines the form of the curve; the simpler the task, the greater the coefficient of precision.

9. A precise measure of speed, when accuracy varies, is suggested.

(Manuscript received November 3, 1934)

References

1. Boring, E. G., Urban's tables and the method of constant stimuli, *Amer. J. Psychol.*, 1917, **28**, 280–293.
2. Boring, E. G., A chart of the psychometric function, *Amer. J. Psychol.*, 1917, **28**, 465–470.
3. Brown, W. and Thomson, G. H., *Essentials of mental measurement.* Cambridge, 1921, pp. 216.
4. Culler, E., Studies in psychometric theory, *J. Exper. Psychol.*, 1926, **9**, 169–190.
5. Fernberger, S. W., The approximation of actual data to the phi-gamma hypothesis, *Amer. J. Psychol.*, 1923, **34**, 498–500.
6. Fisher, R. A., *Statistical methods for research workers.* 4th Edit., London, 1932.
7. Hoisington, L. B., An example of the fractionation of data from the method of constant stimuli for the two point limen, *Amer. J. Psychol.*, 1917, **28**, 588–596.
8. Kelley, T. L., *Statistical method.* N. Y.: Macmillan, 1923, pp. xi + 390.
9. Lanier, L. H., The interrelations of speed of reaction measurements, *J. Exper. Psychol.*, 1934, **17**, 371–400.
10. Lufkin, W. M., The best fitting function for Urban's lifted weight results, *Amer. J. Psychol.*, 1928, **40**, 75–82.
11. Newhall, S. B., Linear interpolation vs. the constant process, *Amer. J. Psychol.*, 1927, **38**, 390–402.
12. Newhall, S. B., An interpolation procedure for calculating thresholds, *Psychol. Rev.*, 1928, **35**, 46–67.
13. Pearson, K. A., *Tables for statisticians and biometricians.* 2nd Edit., Cambridge, 1924.
14. Rich, G. B., Psychophysical measurement methods, *Psychol. Bull.*, 1925, **22**, 613–648.
15. Thomson, G. H., A direct deduction of the constant process used in the method of right and wrong cases, *Psychol. Rev.* 1919, **26**, 454–464.
16. Titchener, E. B., *Experimental psychology.* Quantitative Manual, 1905.

MOTOR CONCOMITANTS OF THE ASSOCIATION REACTION

BY HAROLD E. BURTT

Ohio State University

Introduction

Some 20 years ago Nunzberg,[1] while his subjects were performing the association reaction experiment, had their hands attached to the Somer apparatus which analyzes the movements into 3 directional components and records them separately. More recently Luria[2] had the subject's finger resting on a pneumatic key which he was instructed to press simultaneously with his verbal response and which recorded the pressure continuously. Luria reports that this instruction was so vivid that when the subject thought of a word without speaking, he pressed the key. This tendency was noted in some aphasics, some murderers and in a controlled association experiment. In further laboratory studies and also with criminals Luria notes a larger percent of 'disturbed reactions' on crucial words. He comments incidentally on tremors in the pneumatic record and on manual responses that were premature or delayed.

The present experiment used essentially the foregoing technique but attempted a more exhaustive quantitative and qualitative analysis of the records rather than a mere determination of the percent of cases in which disturbances were found. It seemed probable that individuals would differ in the actual shape or pattern of the curve traced by the marker which recorded from the pneumatic key and also in the types of divergence from this individual pattern which were diag-

[1] *Cf.* Lowenstein, O., *Experimentelle Hysterielehre*, 1913.

[2] Luria, A. R., Die Methode der abbildenden Motorik bei Komunikation der Systeme und ihre Anwendung auf die Affektspsychologie, *Psychol. Forsch.*, 1929, **12**, 127–174, also Luria, A. R., Die Methode der Abbildenden Motorik in der Tatbestandsdaignostik, *Zeits. f. ange. Psychol.*, 1930, **35**, 139–183.

nostic of emotional disturbances. Furthermore it seemed desirable to perform the initial experiment on a rather mild emotional level in the conventional fashion. If diagnostic criteria could be found under these conditions the presumption is that they would be still more significant when the emotional situation was more acute.

Method

The procedure aside from the pneumatic response was quite conventional. The subject, following typewritten directions, examined one of two boxes,—*A* containing objects pertaining to cards, smoking and liquor, and *B* toilet accessories and cosmetics. The typewritten stimulus words pertaining to *A*, *B* and neither (20 each) were presented in an exposure apparatus and the subject spoke into a voice key. These mechanisms actuated a Dunlap chronoscope for conventional analysis of the verbal association times. In parallel with the chronoscope magnets were signal magnets recording on the polygraph for comparison with the manual response. At the front end of the arm of the subject's chair and level with it was a hinged cover on a box inside which a metal bellows was mounted. The top of the bellows was attached to the hinged cover and the subject's first and second fingers were clamped to the top of this cover. Our procedure at this point differs slightly from Luria's in that the cover came to a stop about one centimeter below the position in which the subject normally held his fingers. Luria employed no stop so that the subject could press as far as he wished. In some preliminary work our subjects were continually inquiring how far they should push the key and it appeared that they were developing some kind of a set toward the apparatus which it was felt might detract from the principal manifestation of the emotional disturbance through the manual response. While the stop made it impossible to use the extent of pressure as a diagnostic device it appeared desirable to incorporate this procedure with a view to avoiding the aforementioned set. The bellows was connected to a tambour which recorded on the polygraph. A pneumograph was likewise used but the results are not included in the present discussion.

In the main series the subject's right hand was used on the pneumatic key. In the light of the findings some data collected earlier by a graduate student[3] were analyzed de novo. This will hereafter be termed the preliminary series. In this case the key was operated with the left hand as the right was resting on a planchette for the study of involuntary movements.

After the adjustment of the apparatus the subject was merely instructed to watch the exposure apparatus after the ready signal and when the word appeared to speak the first word which occurred to him as quickly as possible, and at the same time 'in order to facilitate the recording' to press the pneumatic key with his finger. None of the subjects experienced difficulty in doing this on the normal words after a few practice words.

Results

The first consideration in analyzing the results is to determine the particular manual response patterns that char-

[3] Camp, Glenn, Motor concomitants of the association reaction. M. A. Thesis, Ohio State University, 1933.

acterize the individual subject. Wide individual differences in this respect were found and before one could decide what features were diagnostic in the individual case it was necessary to ascertain the general trend for the subject. For instance a slight tremor before pressing the key might be quite significant of an emotional disturbance in one subject whereas another individual might show that pattern practically all the time.

Inspection of the record usually enabled the experimenter to classify the individual pattern as one of several types. The most frequent of these are included in Fig. 1 along with some other variations to be discussed later. In each case the record reads from left to right, *i.e.*, the left end was temporally

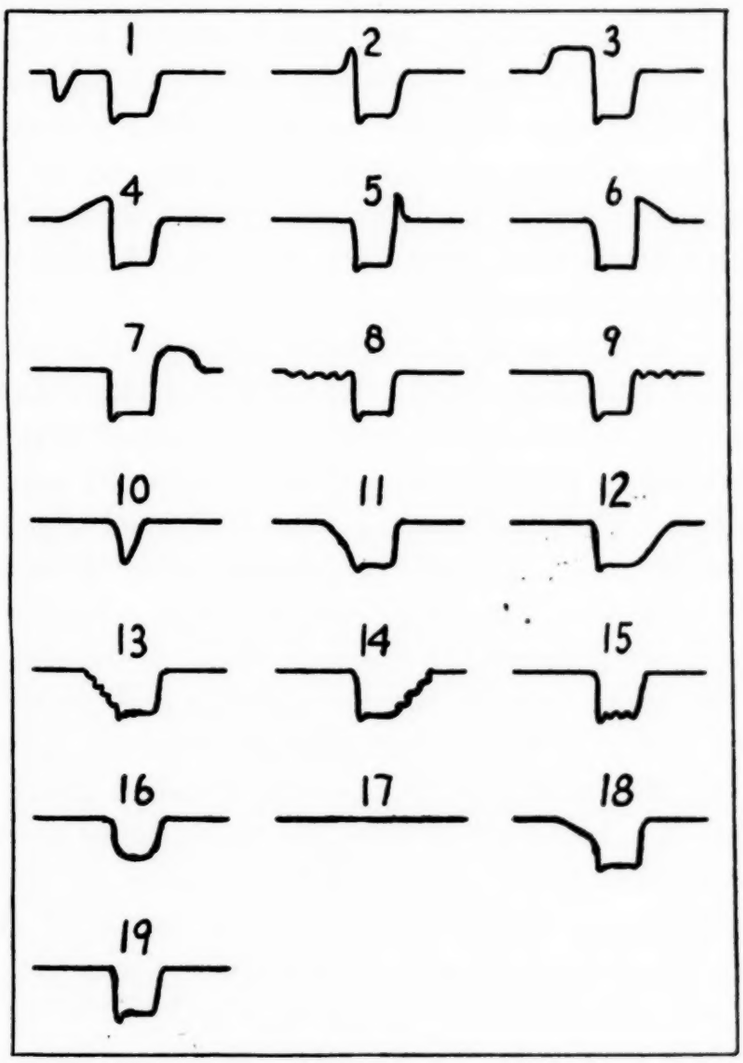

FIG. 1.

earlier, and a downward stroke indicates pressure on the key. The record of stimulus signal magnets, breathing, etc., is omitted in the interest of clarity. At 19 in the figure we have the most frequent pattern. The subject presses the key and holds it for a fraction of a second before letting it up. This 'trough' varies from about 0.2 to 1.0 seconds with the different subjects. This pattern was characteristic for 13 of the 29 subjects who participated. Next in frequency is the variation shown at 5. It is like the preceding except that after the reaction there is a little rebound effect in which the finger is lifted to higher than its initial level and then returns. This type characterizes 8 subjects. In type 10 the subject presses the key and lets it up immediately, both initial and final positions being about the same. This type characterizes 5 subjects. The remaining 2 subjects tended toward a combination of 2 and 5. This involves a slight pull upward before pressing the key and then the rebound effect afterward. A given subject did not, of course, stick perfectly to his type but his general trend was obvious from inspection of the record, and deviations from this type or 'disturbances' in Luria's sense were easy to locate.

We may now turn to the aspects of the results which may be expressed quantitatively. The first of these has to do with the actual moment at which the manual response occurred relative to the verbal. In most cases the manual was a bit earlier. The discrepancy was measured algebraically in millimeters or, as the polygraph was driven by a synchronous motor, converted into seconds. In the case of each subject we secured the mean of these measurements for each of the three lists of words—normal and pertaining to box *A* and box *B*. Three comparisons between means were possible but the most crucial one was between those for box *A* and box *B*. The results with the non-crucial words were used as supplementary to the interpretation of this most crucial difference. The critical ratio (in terms of standard deviation) was computed and in the subsequent tabulations the cases are classified according to whether the ratio was greater or less than 2.0. Hereinafter the former cases will be

designated as 'significant' although it is realised that the exact point at which a difference becomes significant is debatable.

In the first row of Table 1 are results for the mean time by

TABLE 1

	Main Series				Prelim. Series				Both Series			
	Correct		Incorrect		Correct		Incorrect		Correct		Incorrect	
	Significant	Not Significant	Significant	Not Significant	Significant	Not Significant	Significant	Not Significant	Significant	Not Significant	Significant	Not Significant
Early *M*	3	5		4	3	10		4	6	15		8
Early σ	4	4	2	2	1	11		4	5	15	2	6
Trough *M*		6		4		7		9		13		13
Trough σ	2	2	1	3	1	7		8	3	9	1	11
Type:												
1		3		1		3		1		6		2
2		2	1	2		5		4		7	1	6
3	1					3		4	1	3		4
4		2		5	1	7		2	1	9		7
5		1		1		8	1	1		9	1	2
6		3		1		5		3		8		4
7				1		2		2		2		3
8	1	2		4		8		5	1	10		9
9		5		2		6		7		11		9
10								2				2
11		1		3		2	1	3		3	1	6
12		1				2		1		3		1
13				1		1		3		1		4
14		2		1		3		2		5		3
15		1								1		
16		2		1						2		1
17		3		1		2				5		1
18		1				1				2		

which the manual response precedes the verbal. The block of the table at the left gives data for the 12 subjects in the main series, the central block for the 17 subjects in the preliminary series and the right block for the total of both series. Successive columns in each block indicate first the number of subjects for whom the average is significantly greater on words pertaining to the box examined, *i.e.*, 'diagnosis correct'; second, the number of cases where diagnosis is correct but the difference insignificant statistically; third, the number of subjects for whom the average is significantly greater on

words pertaining to the box not examined, *i.e.*, 'diagnosis incorrect'; fourth, the number incorrect but by an insignificant difference. The cases in which the means or other measures to be described below are equal, *i.e.*, no difference at all, are omitted from tabulation. We note 3 subjects in the preliminary series and 3 in the main series for whom this particular variable is diagnostic by a statistically significant amount while there are no significant differences in the other direction. If we include also the less significant differences of 5 and 10 respectively, we have a distinct predominance of correct over incorrect diagnosis. It appears that with quite a number of individuals the extent to which they press the key ahead of the voice reveals the emotional disturbances involved in the association experiment.

It is somewhat the same story with reference to the standard deviation of the measurements just discussed. Data appear in the second row. In two cases the results point significantly in the wrong direction. Otherwise the predominance of correct diagnosis is about the same as when using the means.

Of somewhat less diagnostic value is the length of the trough. If we refer to curve 19 in Fig. 1, the length of the horizontal portion at the bottom indicates how long the subject held the key down. Some subjects, of course, did not manifest such a trough at all. The third row of Table 1 indicates that the mean of these trough lengths is a bit diagnostic in the main series although somewhat negative in the preliminary series. With reference to the standard deviation of these lengths (fourth row) the statistically significant differences are more often in the correct direction but otherwise the results are unimpressive.

Certain other quantitative factors were investigated in individual records but did not occur with sufficient uniformity to be analyzed with reference to the group as a whole. In curves like no. 18 (Fig. 1), the time from the beginning of the gradual drop until the place where the abrupt drop occurred was of interest in a few cases. In curves like no. 3, the time by which the initial rise preceded the main reaction was of

some significance. In some instances the actual amplitude of the push was measured but usually the subjects pressed the key to a maximum and came back to about the initial level. With a few curves like no. 5 the height of the rebound above the final level was determined. The foregoing are the only quantitative aspects that seem worth further study in subsequent records.

Turning to the qualitative features of the records, some 30 different variations were observed wherein the curves for a subject might deviate from his normal record. Some of these are shown in Fig. 1. The most frequent normal types have already been noted but any one of them might represent a variation for another subject. The different variations will be explained and then the frequency of some of them discussed in more detail.

The greatest interest attaches perhaps to no. 1—a double response. The subject presses prematurely and then makes a second reaction more nearly synchronous with the verbal response. In no. 2 we have a sudden rise before pressing the key. In no. 3 this sudden rise occurs somewhat in advance. In no. 4 there is a gradual rise before pressing. In no. 5 there is a rise after pressing to a level higher than the initial and then a sudden recovery to the initial level. In no. 6 this recovery to normal is gradual. In no. 7 this recovery does not come down uniformly in a straight oblique line but is somewhat rounded. No. 8 involves irregularities or tremors before the response and no. 9 irregularities after the response. In no. 10 the hand comes up instantly without a trough. In no. 11 the pressure is applied gradually rather than with the usual abruptness while in no. 12 the finger is lifted somewhat gradually after the reaction. No. 13 involves a sort of staircase effect on the way down while no. 14 involves a similar effect on the up-pull. No. 15 involves irregularities in the trough and no. 16 is a sort of rounded trough. No. 17 is the omission of the manual response altogether. In no. 18 the subject starts to press gradually and then does so suddenly. No. 19 is a normal type already discussed. Other variations which occurred in the case of only a single subject are not included in the figure.

The lower part of Table 1 summarizes results for the qualitative variations just discussed. The numbers in the left column correspond to the numbers of the curves in Fig. 1. The remaining columns have the same meaning as in the upper part of the table. With reference to the statistical analysis of these data one point should be noted. The comparisons indicated in the table are based on computation of the percent of the words in list *A* on which a particular variation occurred as compared with the percent in list *B* manifesting that same variation. In each case we determined the standard deviation of the difference between the two percents and computed the critical ratio. However, it is quite conceivable that the emotional aspect for the box examined might be pronounced in only a few of the 20 words in a given list so that a given type of disturbance would occur only a few times but it might not occur at all in the other list. From a statistical standpoint this would mean an insignificant difference but from the standpoint of diagnosis a few disturbances in one list and none in the other might be of considerable interest. The scarcity of entries in the 'significant' column in this portion of the table should be interpreted in the light of the foregoing suggestions.

We will not discuss the rows in the qualitative portion of the table seriatim but merely note the features which appear most diagnostic. The double response (no. 1) is the disturbance which Luria stressed in his analysis. In our experiment with rather unexciting situations, these double responses were found in the results of only 4 individuals in each series. However in 6 of the 8 cases the double response occurred more frequently on guilty words. We experimented with other methods of analysis which might be valuable if these responses occurred with greater frequency. We might, for instance take the ratio of the amplitude of the first response to the second on the theory that a greater emotional disturbance would produce a comparatively more vigorous premature response. Or we might measure the time interval by which the premature response preceded the final. In the 4 cases in the main series both these methods yielded results

that pointed in the same direction as those included in the table.

Of the remaining disturbances in the main series the most diagnostic are 9, 6, 17, and 3. The last is based on only one subject but it occurs with significant frequency in the guilty list. The results for the preliminary series corroborate the main one on item 17 but are somewhat equivocal on the other three. Three further items are of interest in the preliminary series only, viz. 4, 5, and 8. If we combine results of both series it appears to be items 5, 6 and 17 that are most diagnostic on the whole. These involve the rebound effect after pressing the key with either sudden or gradual return to initial level and the omission of the response altogether.

If we total the columns for the qualitative results in the table we get somewhat larger totals in the 'correct' columns, indicating that the emotional disturbance on guilty words manifests itself in general by more frequent irregularities in the record. However it is of more interest to note the types of irregularity that are most frequently diagnostic.

Another analysis of the data may be made from the standpoint of the individual subject, following essentially the procedure that would be used in a practical situation. We determine for a given subject the types of irregularity or disturbance that predominate in list *A* and in list *B*. If several of these show greater frequency in list *A* and none predominate in list *B*, the presumption is that *A* is the 'guilty' list. To show the extent to which this procedure would be successful, Table 2 has been prepared.

Each row represents an individual subject. The upper block is for the main series and the lower for the preliminary series. The entries indicate the number of items or variations out of the 22 listed in Table 1 that occurred for a given subject more frequently in one list than the other. As before, they are classified according to whether the difference between means, sigmas or percents is greater or less than twice the standard deviation of the difference. At the right these entries are totalled disregarding significance. For example, with subject A there were 6 items that pointed in the correct

TABLE 2

Subject	Correct		Incorrect		Total	
	Significant	Not Significant	Significant	Not Significant	Correct	Incorrect
A	0	6	0	2	6	2
B	0	4	0	0	4	0
C	3	1	1	4	4	5
D	2	1	0	4	3	4
E	1	6	0	2	7	2
F	0	5	0	2	5	2
G	1	3	2	1	4	3
H	0	3	0	7	3	7
I	1	5	0	1	6	1
J	0	1	0	6	1	6
K	0	7	1	5	7	6
L	3	3	0	3	6	3
Preliminary						
M	0	3	0	5	3	5
N	1	6	0	3	7	3
O	0	4	0	6	4	6
P	0	8	0	4	8	4
Q	0	6	0	6	6	6
R	0	2	0	2	2	2
S	1	4	0	4	5	4
T	1	8	0	1	9	1
U	0	5	0	3	5	3
V	0	8	0	2	8	2
W	0	3	1	6	3	7
X	1	3	1	1	4	2
Y	0	10	0	2	10	2
Z	1	3	0	4	4	4
AA	0	8	0	8	8	8
BB	1	5	0	4	6	4
CC	0	7	0	4	7	4

direction and 2 in the reverse but no differences were statistically significant. If we had been compelled to make a diagnosis, however, we would have followed the 6 rather than the 2 and made a correct judgment. On a similar basis we would have judged correctly in 8 of the 12 subjects in the main series. Furthermore with subjects C and D the totals are pretty close but the significant differences point rather clearly in the correct direction so that we would presumably have been correct with 10 of the 12 subjects.

With the preliminary series in the lower part of the table our success would be less pronounced. On the basis of totals, 10 of the 17 would be judged correctly. There were three other cases (Q, R and AA) where it would be a sheer guess

because totals in both columns were equal. In another such case (Z) judgment would probably be correct because of one significant difference in the correct direction.

By way of comparison with our results the conventional analysis was made of the association reaction times, computing the difference between means of lists *A* and *B* and also between sigmas. In the main series only 4 of the differences between means pointed in the correct direction and none of these were significant. Seven of the differences between sigmas were correct and one of them significant. In this main series our analysis of the manual reaction was obviously more successful than the usual analysis of the association reaction times. In the preliminary series 11 of the differences between means of association time were correct but none significant. Nine of the differences between sigmas were correct and 3 of them significant.

We feel that the experiment indicates a number of things regarding future work. The right hand seems preferable for the manual response (we had no left-handed subjects). Our main series was on the whole more successful than the preliminary. While some of this may have been due to a more experienced experimenter it is possible that another factor was the use of the right hand in the main series as contrasted with the left in the preliminary. Some other preliminary work with the right hand but a poor recording technique strengthens this conviction. It is plausible that the subject who had used the right hand in acts of skill would have a stronger manual set with that hand and be more apt to make premature reactions or at least do something diagnostic with that hand.

Again it would seem that we have developed essentially the procedure of analyzing the records that could be used profitably in further work for theoretical or practical purposes. The steps in this procedure have been outlined: (1) inspect the record as a whole to get a notion of the general pattern of the subject; (2) annotate all reactions on the record that differ qualitatively from the subject's general pattern. It is probable that the 18 listed in the figure are about all that may be

expected to be of interest. A convenient code may be used for these annotations. They should be made before the reactions are marked according to the actual stimulus word involved, thus eliminating any element of suggestion on the part of the experimenter in deciding whether a reaction is sufficiently different from the general pattern to be noted. (3) Measure the time by which the manual preceded or followed the verbal reaction and the length of the trough. (4) Mark the reactions to show the stimulus word and then assemble all items that have been annotated, in terms of crucial or non-crucial words. Critical ratios should likewise be computed. Then if most of the variations noted occur with greater frequency and especially with statistically significant frequency in the crucial list we are in a position to conclude that these words touched off more emotional disturbance or aroused complexes or indicated guilt, according to the conditions of the experiment.

Finally, it is suggested to the crime detectors that while the association reaction has been considerably overshadowed by blood pressure, etc., it would be profitable to give it another 'try' using this supplementary manual technique.

Summary

The experiment followed Luria's technique, having the subject press a pneumatic key at the same time he gave his verbal response. This key recorded continuously the pressure exerted by the hand. More exhaustive quantitative and qualitative analysis of the records was attempted in an experiment which was otherwise conventional involving the determination of which of two boxes a subject had inspected. Subjects differed in the typical patterns traced by the recorder from the pneumatic key. Most of these normal patterns fell into five types. It was then possible to note significant disturbances or deviations from this normal pattern for a given subject.

Quantitatively the most diagnostic feature was the time by which the manual response preceded the verbal though it was supposed to be synchronous. The standard deviation

of these times was likewise of some diagnostic value. In the case of subjects who pressed the key and held it down for a fraction of a second before releasing, the length of this period was somewhat significant in differentiating which box they had examined and of less significance was the standard deviation of this length. The qualitative variations comprised 18 which were of sufficient frequency to be useful in subsequent analysis. A double response in which the subject presses prematurely and then presses again at the proper time was perhaps the most interesting of these. The other variations which are most frequently significant are the rebound after pressing the key with either a sudden or gradual return to initial level and the omission of the manual response altogether. From the general standpoint of diagnostic value the present technique proved better than the conventional analysis of association-reaction time. A suggested procedure for routine interpretation of such records is presented.

(Manuscript received October 31, 1934)

HANDEDNESS EFFECT IN SIMULTANEOUS LIFTING OF WEIGHTS BY BOTH HANDS

BY N. C. SHEN[1]

Psychological Laboratory
Tsing Hua University
Peiping, China

In a former experiment on *Simultaneous lifting of equally heavy weights by both right and left hands* (1), the writer found that all the subjects, being right-handed, tended to pass many more 'right-heavier' or 'left-heavier' judgments, although theoretically these two kinds of judgments, if passed at all, should be equal in number. This preponderance of right-heavier or left-heavier judgments may be attributed to the effect of handedness, but the fact that some gave many more right-heavier and the rest many more left-heavier judgments makes the handedness explanation untenable, because all the subjects are right-handed. Besides, other sources of check seemed to indicate that handedness did not play an appreciable part in determining the type of preponderant judgments in one category or the other; instead, 'attitude' was considered to be rather powerful in modifying the judgment trend.

As the original problem of the former experiment was to investigate, if possible, the role of handedness in lifting weights, the outcome that it is attitude rather than handedness that had definitely influenced the trend of judgment means a serious failure to meet the expectation. Whether this failure was due to a wrong procedure or to the kind of subjects used, there was no indication to settle the question; and this difficulty necessitated the undertaking of the present experiment, in which we chose arbitrarily to vary the type of procedure.

[1] The writer wishes to acknowledge his indebtedness to Prof. Siegen K. Chou for his general direction and constructive criticisms. Sincere thanks are due to the ten subjects for their willing coöperation and untiring interest in the experiment.

Problem

In simultaneous lifting of weights by both the right and the left hands, inequality in judgment tendency peculiar to one hand may be due to the manifestation of handedness effect. In saying that, we have in mind, as a natural result of handedness, an inequality in the capacity of exerting force by the preferred and the non-preferred hands. As the perception of heaviness or resistance depends partly on the way of lifting, *i.e.*, the amount of force exerted, it is very conceivable to expect a difference in judging simultaneously lifted weights by both hands, one of which being the preferred. To be more exact, the preferred hand, when matching a weight psychologically equal in heaviness to that lifted by the non-preferred hand, the former should choose a heavier weight than does the latter, and vice versa. The assumption is that, the preferred hand, being capable of exerting greater force, should under-estimate a heavier (than the one lifted by the other hand) weight and regard it as equal to the one lifted by the non-preferred hand.

Now the problem of the present experiment is to study the kind of judgments passed by the preferred hand of each subject when both hands are required to lift weights simultaneously, one in each; and one of the hands is to choose a match perceived to be equal in heaviness to the standard weight in the other. In other words, with simultaneous lifting by both hands, we aim to find out from the perceived-to-be-equally-heavy weights peculiarities in judgments by the preferred hand.

Method and Procedure

1. Subjects and Preliminary Practice.—Ten adult right-handed men served as subjects. Six of them are advanced students in Psychology and four assistants in the Department. Each subject was blindfolded and given a preliminary practice on simultaneous lifting of weights, one hand having the 'standard' weight and the other selecting an equally heavy 'variable' weight to match the former. This preliminary practice consisted of twelve judgments in all, six times with the standard weight in the right and six times with it in the left hand. The standard was 100 g. throughout. The matched variable weight was recorded each time, but the subject did not know its actual weight.

2. The Weight and the Technique of Lifting.—There are two kinds of weights, the standard and the variable. All the weights are round wooden boxes 2½ in. in diameter

and 2½ in. in height and painted in black. We used one standard, 100 g., and sixteen variable weights, ranged from 79 to 124 g. by steps of three. The variable weights were arranged on the periphery of a rotating platform, 20 in. in diameter, with a space of about ½ in. in between any two weights. For each weight on this platform there was a depression ¼ in. deep to accommodate it. Just by each side of the rotating platform was a support for the standard weight, being of the same height and with a similar depression. When the right hand was presented with the standard weight, the support on the left side was not used, and vice versa.

When the blindfolded subject was seated comfortably and his arms adjusted on the cushions which could be moved along an axis to fit his breast width, he was to use the thumb and the middle finger of each hand to lift the weights in the middle. For example, if the standard weight is presented to the right hand, the left hand is to pick up an equally heavy variable weight from a definite end of the series, which is either in an ascending or a descending order. The employment of ascending orders in alternation with descending orders is to get rid of expectation on the part of the subject and to set a partial control over the effect of a preceding weight upon the judgment of the later one as demonstrated by Guilford and Park (2) and Turner (3).

The subject was required to lift up the standard and the variable *at the same time to approximately the same height*, and *lower down the two kinds of weights at the same time*. This procedure of lifting was repeated for every variable weight picked up for comparison until the right variable was reached.

For different trials, the subject started from different parts of the variable weight series. This was done to avoid suggestion and guessing. Altogether three 'sizes' of the variable weight series were employed, namely 'short,' from 88 upward or from 112 downward; 'intermediate,' from 82 upward or from 118 downward; and 'long,' from 79 upward or from 124 downward. In addition to that, ascending and descending series alternated in order to avoid the factor of expectation in the method of limits.

3. *Instruction.*—The instruction was typed on a sheet and given to each subject when he came for his first sitting. Regular liftings were not started until the subject had no question as to what he was supposed to do. The instruction reads as follows:

"I am going to present a box of weight to one of your hands, please use your other hand to select an equally heavy box from the series put on the round platform. This series of boxes is arranged in order of heaviness. When I move the platform, the boxes arranged on it will be either increasingly heavier or increasingly lighter.

"In trying to pick out the box of weight which appears to you to be equal in heaviness to the one which I present to you, please go over the series of boxes one by one carefully until you get the right box. When you have got the right box, just say 'here' and stop selecting any further.

"Please judge very carefully every box which you pick up for comparison. Sometimes you need only pick up a few boxes of weight before you get the right one. Sometimes you have to pick up a great number of them before you get the right one. Just do your best and your judgment will be most likely correct!"

4. *Spatial and Temporal Error.*—Our procedure required every subject to lift the standard and the variable weights simultaneously, so temporal error is practically out of question. There might, however, be something similar to temporal error working only in the variable weights, that is, the correct variable weight regarded as equally heavy to the standard was generally not reached until after going over quite a few steps from a certain end. Thus the matched variable weight might be either too much heavier or too much lighter than it should be if the subject did not have to go over the

series in succession. This possible error was controlled by alternating ascending with descending series, for Turner had found that the modifying effect of a heavier preceding weight upon the judgment of a later one was opposite to that of a lighter preceding weight.

As to spatial error, the schedule for each sitting provided equal number of judgments for the right hand and the left hand to lift the standard. Furthermore, this experiment, similar to the former one by the same writer, may be considered as studying the very spatial error in lifting weights with reference to the two hands. When other conditions have all been equated for the two hands, what comes out to be different for them may be attributed to the operation of handedness. In other words, the manifestation of handedness difference in lifting weights amounts to the same thing as the expression of spatial error here.

5. *Schedule of Lifting.*—Each subject had a preliminary practice as mentioned above. After that, generally from the following day on, he went through two sittings of 40 judgments each, one in a day at the corresponding hour. A sitting consisted of 20 judgments each with the standard weight in the right hand and the left hand. The order of presenting the standard to one of the hands was like the following:

A		B	
R hand	10 times	L hand	10 times
L "	" "	R "	" "
L "	" "	R "	" "
R "	" "	L "	" "

If schedule A be taken for the first sitting, schedule B serves the second, and vice versa.

After every ten judgments, that is, before the change of the standard weight from one hand to the other, a rest of about two minutes was allowed. This interval was not held strict, and generally a longer one was given if the subject complained of fatigue.

Results

As our procedure calls for the selection by one hand of an equally heavy weight to match the standard in the other, there are two possibilities in the results if the two hands were not different by any appreciable degree: the average of the variable weights selected in different trials by either the right or the left hand to match the standard should approximate 100 g.; or the average variable weight chosen by the right or the left hand is consistently less or more in magnitude than the standard weight. In the latter case, a constant error of either over-estimation or under-estimation of the variable weight seems to be at work, but this result still means an equality between the two hands. On the contrary, if the resultant average variable weight should be consistently smaller than the standard for one hand and consistently

larger for the other, or the difference between the average variable weights lifted by the two hands be consistent and significant, then the two hands seem to be different with reference to lifting weights.

Over- and Under-estimation by Left and Right Hands.— Our results indicate a consistent tendency of over-estimation of lifted weights by the left and under-estimation by the right hand, when a standard weight was used and the technique of simultaneous lifting by both hands employed. When the left hand received the standard weight, the average variable weight chosen by the right hand to match the former as psychologically equally heavy was larger than the other way, that is, with the right hand lifting the standard and the left hand selecting the variable weight. Table I gives for the ten subjects the average variable weights chosen as equally heavy as the 100 g. standard. The table treats the two hands separately and includes the results from the preliminary practice.

TABLE I

AVERAGE VARIABLE WEIGHTS JUDGED AS EQUALLY HEAVY AS THE STANDARD WEIGHT (100 G.)

Subject	Standard Weight in Right Hand		Standard Weight in Left Hand	
	(A)	(B)	(A)	(B)
1	95.5	97.0±.61	106.3	106.9±.59
2	97.0	95.7±.68	104.0	103.7±.59
3	91.0	91.4±.61	111.8	108.3±.77
4	96.0	97.9±.39	103.5	103.7±.47
5	100.5	97.8±.43	103.5	102.0±.35
6	98.5	97.4±.48	101.0	102.8±.43
7	99.0	98.6±.46	103.5	102.8±.38
8	99.0	97.9±.62	102.5	102.8±.54
9	100.0	97.1±.70	91.0	98.9±.44
10	98.5	104.6±.62 (97.3±.65)*	99.0	100.5±.55 (103.6±.64)*

(A) Preliminary practice, each figure based upon 6 judgments.
(B) Experiment proper, each figure based upon 40 judgments.
* Results from an extra check experiment.

Besides indicating a clear tendency of over-estimation by the left hand and under-estimation by the right hand, Table I shows that this tendency has already been manifested even

in the few preliminary practice judgments. The only exception to the consistent tendency of over- and under-estimation as mentioned above is found in subject 10 (under column B), who has exactly the opposite tendency. A check on his introspective data revealed, however, that the reversion was due to some accidental illusion during lifting. He reported to have experienced the variable weight as larger (in size) than the standard when the former was lifted by the right hand. Usnadze's experiment (4) had demonstrated the existence of the inverse of the size-weight illusion, *i.e.*, with blindfolded subjects, the heavier of two equally large objects seemed smaller. Thus subject 10's case might well be due to the operation of such an illusion. So, without informing him of anything, an extra check experiment of 40 judgments was undertaken forty days afterwards. In that extra sitting, he was kept unknown as to the purpose of the experiment, but simply told to pay attention only to the felt heaviness of the weights lifted. The results thereof appear in parentheses. It is of interest to note that when the illusion, the inverse of the size-weight illusion, was supposedly absent, his results came in line with those of all the other subjects.

Reliability of the Differences between Over- and Under-estimation.—We have seen above that all the subjects over-estimated and under-estimated the weights lifted by the left and the right hands respectively. The question now before us is to ask whether the difference between over- and under-estimation for each subject is statistically reliable. For that we have computed the "critical ratios' with the formula D_m/P.E.$_D$ as shown in Table II. The conventional practice of taking a critical ratio of 4 or above as meaning practical certainty for having such a difference is well fulfilled by our results, except with subject 9, in whose case the critical ratio is only 2.2. The smallness of this ratio is not due to a large value of the P.E. difference, but due to the smallness of the mean difference between the two kinds of average variable weights as can be seen from the table. We have, however, good reasons to believe, from his introspective report, that the small mean difference resulted from some uncontrolled

experimental factor. The subject complained, during the experiment, of fatigue with his right hand when the left was normal, and of greater fatigue with the right when both hands were fatigued. His complaint seemed only too natural as he had always one hour's hard work of writing with the right hand before coming to the experiment. This greater fatigability of his right hand might have caused a diminished capacity of that hand to exert force in lifting and consequently obliterated, so to speak, the difference between the right and the left hands, so far as the general tendency of over- and under-estimation was concerned.

TABLE II

DEGREE OF SIGNIFICANCE OF THE DIFFERENCE BETWEEN RIGHT-HAND UNDER-ESTIMATION AND LEFT-HAND OVER-ESTIMATION

Subject	Mean Difference Between Over- and Under-estimation	P.E. Difference	Critical Ratio
1	9.9	.85	11.7
2	8.0	.90	8.9
3	16.9	.98	17.2
4	5.8	.61	9.5
5	4.2	.56	7.5
6	5.4	.65	8.3
7	4.2	.59	7.1
8	4.9	.82	6.2
9	1.8	.82	2.2
10	6.3	.91	6.9

Effect of 'Expectation.'—One of the most serious difficulties with the method of limits is the operation of 'expectation.' The adoption of the ascending series of variable weights in alternation with the descending ones aims at balancing expectation working from two opposite ends. That expectation has clearly influenced our results could be seen from Table III where the average variable weights chosen in both ascending and descending series as equally heavy as the standard were given. The table also treats the right- and the left-hand lifted variable weights separately.

Table III tells us not merely the great influence of expectation on the total results, it also reveals how the two hands could be different even under the same condition, that is,

either in ascending or in descending series. In no case was a left-hand lifted variable weight heavier than a right-hand lifted one, except with the descending series of subject 9, where the two figures are practically equal (99.9 : 100).

TABLE III

AVERAGE VARIABLE WEIGHTS JUDGED AS EQUALLY HEAVY AS THE STANDARD WEIGHT IN ASCENDING AND DESCENDING SERIES

Subject	Ascending			Descending		
	R-hand	L-hand	Total	R-hand	L-hand	Total
1	106.3	93.9	100.1	107.5	100.0	103.8
2	100.9	90.6	95.8	106.4	100.8	103.6
3	106.2	90.0	98.1	110.5	92.8	101.7
4	101.5	96.0	98.8	105.9	99.8	102.9
5	101.7	96.9	99.3	102.3	98.7	100.5
6	101.4	95.5	98.5	104.2	99.1	101.7
7	101.7	95.8	98.8	103.9	101.4	102.7
8	99.4	93.1	96.3	106.0	102.6	104.3
9	97.9	94.2	96.1	99.9	100.0	100.0
10*	102.4	94.6	98.5	104.8	99.7	102.3

*Using results from the check experiment.

One question may, however, be asked here. Could this factor of expectation be merely an indication of the naïveté of our subjects? If it be so, we ought to find the results from the second sitting approximating more closely to 100 than those from the first sitting. As Table IV shows, it is, however,

TABLE IV

AVERAGE VARIABLE WEIGHTS JUDGED AS EQUALLY HEAVY AS THE STANDARD IN THE 1ST AND THE 2ND SITTINGS

Subject	Ascending		Descending	
	1st Sitting	2nd Sitting	1st Sitting	2nd Sitting
1	101.2	99.0	104.7	102.9
2	95.2	96.3	102.7	104.5
3	97.3	98.8	100.5	102.9
4	98.1	99.4	102.9	102.9
5	98.1	100.5	101.4	99.6
6	99.1	97.8	101.5	101.8
7	98.8	98.7	103.2	102.1
8	97.2	95.4	103.6	105.0
9	97.5	94.6	102.9	97.0
10	99.4	99.7	105.3	105.6

not the case at all. No consistent change in the numerical values of the results could be found as indicating a weakening of the influence of expectation. So we have perhaps to consider this great influence of expectation as either inherent to the method of limits or to the particular type of procedure here adopted, or to the combination of both.

Variability of Right-hand and Left-hand Lifted Variable Weights.—We have found that, under the same conditions, the right-hand lifted variable weights are definitely physically heavier than the left-hand lifted ones, when the procedure requires the matching of psychological equal heaviness. That finding is sufficient to bring out the difference of the two hands in simultaneous lifting of weights. But handedness difference could be viewed from another angle, *i.e.*, from the point of view of accuracy of judgment. For that, Table V which represents the variability of judgment serves to give some information on the question.

TABLE V

STANDARD DEVIATIONS OF THE RIGHT-HAND AND THE LEFT-HAND LIFTED VARIABLE WEIGHTS

Subject	R-hand Lifted Weights	L-hand Lifted Weights
1	5.61	5.82
2	5.55	6.42
3	7.26	5.82
4	4.50	3.69
5	3.27	4.11
6	4.11	4.59
7	3.63	4.38
8	5.10	5.91
9	4.17	6.60
10	5.22 (4.29)*	5.85 (4.41)*

* Check experiment.

The general tendency seems to favor the right-hand lifted weights with a smaller variability. In eight of the ten cases, the standard deviation of the judgments is smaller with the right-hand lifted weights. It is quite conceivable that a smaller variability means a higher degree of accuracy, and this assumption seems all the more reasonable in view of the fact that the non-preferred hand of the subjects is the left one.

Practice Effect on Handedness Difference in Lifting Weights. —The found difference between the right and the left hands in making a match to the standard weight might not be a genuine one, and the great magnitude of this difference is possibly due to some exaggeration of a lack of practice on the part of the subjects. Though we had no experiment to check on this point, the results permit, however, a general analysis to answer the suspicion. The results for the first sitting are treated separately from those for the second as in Table VI which shows that the latter are not appreciably different from the former so far as over- and under-estimation by the left and the right hands are concerned.

TABLE VI

AVERAGE VARIABLE WEIGHTS IN THE FIRST AND THE SECOND SITTINGS

Subject	R-hand Lifted Weights		L-hand Lifted Weights	
	1st Sitting	2nd Sitting	1st Sitting	2nd Sitting
1	109.9	103.9	96.0	97.9
2	104.4	103.0	93.6	97.8
3	108.9	107.8	88.9	93.9
4	102.3	105.1	98.7	97.2
5	101.7	102.3	97.8	97.8
6	103.2	102.4	97.5	97.2
7	102.9	102.7	99.1	98.1
8	102.6	102.9	98.2	97.5
9	99.0	98.8	101.4	92.8
10	100.2	100.8	104.5	104.7
Average	103.5	103.0	97.6	97.5

If the great difference between the two hands in judging be due to a lack of practice, the second sitting ought to yield a small difference than does the first. In other words, with more practice, for both right-hand and left-hand liftings of the variable weights, the results should tend to approach more closely to 100 g. As Table VI stands, there is no evidence for such a change in the second sitting as compared with the first. Furthermore, in the last row of the table, the averages of the first sitting and those of the second are so alike that the most reasonable explanation would be to

consider practice as negligible in modifying the influence of handedness difference in lifting weights.

Discussion

The outcome of the present experiment has shown definitely that the right hand of the subjects, all being right-handed, as compared with the left, under-estimates lifted weights. Expressing in another way, if the right and the left hands of a right-handed person be used in lifting simultaneously a pair of physically equally heavy weights, the former feels the weight in it as lighter than that in the latter. Since this phenomenon of under-estimation of heaviness by the preferred hand has been found to hold good for all the subjects studied, a tentative explanation should perhaps be sought in terms of handedness difference.

The preferred hand, as it is used much more frequently in daily life, becomes naturally more capable of exerting greater force than does the non-preferred in reacting to resistance, when the conditions for the two hands are held equal. Now, as the time of lifting is simultaneous, the size of the weights same, and other environmental circumstances equated for the two hands, the feeling of 'lighter-ness' of the weight lifted by the right hand seems simply to be a different expression of the size-weight illusion, where the fact that the larger one of a pair of equally heavy weights is judged as lighter is because of the intention of the subject to exert greater force in lifting the larger sized weight. When size is made equal, the size-weight illusion of course vanishes, but the inherent capacity of the preferred hand to exert more energy amounts to substitute the factor of size with handedness and consequently makes the weight lifted by this hand feel lighter than that lifted by the non-preferred hand.

Besides this exerting-greater-force capacity, the preferred hand is likely to be the one that experiences heaviness more accurately. The results that, in most cases, variability of judgments is smaller with the right-hand lifted weights than with the left-hand lifted ones could be interpreted in this way. This higher degree of accuracy of lifting weights by

the preferred hand may well be due to frequent training of it through innumerable opportunities of weight-estimation that come in contact with our hands but generally go to the share of the preferred hand. Thus the lifting weight experiment affords a new technique of testing some aspects of the phenomenon of handedness which have hitherto been studied by non-psychological means.

Finally, a word may be said for the limitations of the application of the present results, as well as for future work on the same topic. As all the subjects used are right-handed, a repetition of our work with left-handed subjects seems extremely important in order to check on the reliability of our results. It is, however, not right to expect just the opposite kind of data with left-handed people, because under the present social organization, the right hand of the right-handed is truly the preferred hand while the left hand of the left-handed may not necessarily be the preferred hand. This statement is not paradoxial at all, if one should take note of the fact that the custom goes with the right hand and ridicules the use of the left hand in all sorts of daily activities. Nevertheless, the left-handed people will perhaps give quite a different story, most probably yielding irregular results as compared with our regular ones.

Summary and Conclusions

An experiment was performed with ten right-handed subjects to study the effect of handedness upon lifting weights. The method of limits was employed and the technique of simultaneous lifting by both hands adopted. The results show that:

1. There was a tendency for right-hand under-estimation and left-hand over-estimation of heaviness. In other words, for the weights lifted by the right hand and the left hand as equal in heaviness to a constant standard of 100 g., the right-hand lifted ones are definitely physically heavier than the left-hand lifted ones. This difference between the two hands was attributed to the supposed capacity of the preferred hand, the right, to exert greater force than does the non-preferred under identical conditions.

2. The difference between the two hands as manifested in lifting weights is statistically reliable in nine out of the ten subjects used.

3. The factor of expectation worked significantly in the results. It is dependent on the direction of the variable weight series, either ascending or descending. But handedness difference was not masked by expectation at all.

4. The right hand, the preferred one, was not only capable of exerting greater force as evidenced by its under-estimation of weights, it was also more accurate than the left hand in making judgments. Here higher accuracy means smaller variability of the results.

5. Practice did not seem to have appreciable effect in modifying the influence of handedness difference in lifting weights.

(Manuscript received October 22, 1934)

References

1. Shen, N. C., Simultaneous lifting of equally heavy weights by both right and left hands, *J. Exper. Psychol.*, 1935, **18**.
2. Guilford, J. P. and Park, D. G., The effect of inerpolated weights upon comparative judgments, *Amer. J. Psychol.*, 1931, **43**, 589–99.
3. Turner, W. D., Intra-serial effect with lifted weights, *Amer. J. Psychol.*, 1931, **43**, 1–25.
4. Usnadze, D., Über de Gewichstauchung und ihre Analoga, *Psychol. Forsch.*, 1931, **14**, 366–79.

THE CONDITIONED REFLEX AS THE EXPLANATION OF HABIT FORMATION: III. THE OPERATION OF TWO HIGHER-ORDER REACTIONS IN CLOSE SUCCESSION

BY J. M. STEPHENS

Department of Education
The Johns Hopkins University

Explanatory Statement and Summary of the Series

The series of which this study [1] is the third part, has taken on two aspects, in a sense different, and yet capable of being confused. It has seemed wise to differentiate between these aspects at some length before proceeding to the discussion of the third study. It has been most convenient to do this in the way of a very brief summary of the series. The summary includes some of the major results of the series—even anticipating those of the study about to be reported.

The General Nature of the Hypothesis to be Examined

The lengthy general title of the series could be stated as a question, viz., "How adequate is the conditioned reflex theory of learning?." By the conditioned reflex theory of learning (henceforth indicated as the CR hypothesis) we mean the theory that the principles (presumably fairly well understood) which describe or predict what happens in the establishment of the conditioned reflex are capable of describing or predicting the outcomes of other types of learning.

Now one of the (if not *the*) crucial principles used to describe the outcomes of the conditioned reflex is that which attempts to predict the *direction* of the conditioning. In speaking of the direction of the conditioning we refer to the

[1] In this investigation I was greatly aided by Mr. R. Milton Hall, Principal of School No. 40., Baltimore, who provided space and facilities and secured the subjects.

fact that one stimulus (say the sound of a bell) comes to take on the response to the other (say weak acid in the mouth), but that the other (acid) does not come to take on the response to the one (sound of the bell). We say that this matter of direction is one of the most significant facts to be explained. Apart from direction, we would merely have association by contiguity in time, which certainly has not sufficient novelty to justify a new theory of learning. It is the possibility of explaining direction which gives the CR hypothesis value, and hence any principle which explains the direction becomes crucial in that hypothesis. The problem of determining the direction of the conditioning is, incidentally, almost the sole concern of this series of studies.

Almost all the principles used to predict the direction of the conditioning have leaned on the concept of dominance. In the example suggested above, the salivary response to weak acid is considered as more dominant than any vague orienting reactions that the dog may make to the sound of the bell. And the direction of the conditioning is to be attributed to this superior dominance.

The use of the term 'dominance' in this series is perhaps not very fortunate, since it has been used by Razran (**4**) in a very special sense. In Razran's use of the term dominance may include the temporal factor. Hence in his view the opposition between the dominance factor and the temporal factor would be ridiculous. In this series, however, the term dominance is used in a much more general sense. In our sense we could speak of the relative dominance of two reactions even though they had never occurred together. In Razran's sense this would be leaving out one possible aspect of dominance—the temporal factor. Perhaps our use of the term could be described as the non-temporal aspects of dominance, as used by Razran.

The First Aspect of the Examination and its Results

In view of the foregoing outline, one aspect of our examination of the CR hypothesis would naturally consist of an

attempt to see if the principle of dominance which we assume for the time to be so crucial in the explanation of the conditioned reflex, does in fact, describe or predict the outcomes of other types of learning. And in its major sense the series has been such an attempt.

In carrying out that attempt it has been necessary to use some empirical criterion of dominance, and since there is some doubt as to its precise meaning, we have used three different criteria—each study employing one. In the first study (**2**) the more intense of two electric shocks was regarded as initiating the more dominant reflex. In that study dominance (intensity) seemed to have no influence on the direction of the conditioning. In the second study (**3**) a lower-order reaction —(withdrawal from shock) was regarded as more dominant than a higher-order reaction (hitting a target in response to a flash of light). Here the results were directly opposite to the prediction of the CR hypothesis, as the stimulus of the more dominant reaction (as here defined) came to take on the response of the less dominant reaction. In the third study (about to be reported) importance was used as a criterion of dominance. In this case, as in the case of the first study, dominance seemed to have little if anything to do with determining the direction of the conditioning. Consequently it seems that the CR hypothesis, insofar as it depends on the principle of dominance, does not appear to describe or predict the outcomes of other types of learning.

The Second Aspect of the Examination and its Results

It is evident, however, that our examination invites an additional line of approach. Early in the series it appeared that, contrary to previous assumptions, dominance was not the principle which determined the direction of conditioning in the conditioned reflex itself. Both the first article of the series (**2**) and the work of Wendt (**5**) and that of Bernstein (**1**) seem to indicate that the temporal factor is much more important than the factor of dominance in this respect. According to these studies this essential principle (the

temporal factor) would be described somewhat as follows: *When two stimuli are presented in conjunction with one another the one which is presented first will come to take on the response of the one presented second.* Now if this is true there is a sense in which one could still say that learning is conditioning. By such a statement, one *could* mean that the temporal factor which explains the direction of conditioning also explains other learning. And (again in a sense) we might be expected to examine the CR hypothesis as modified above.

I have indicated those additional expectations in anything but categorical terms and that for a reason. While the CR hypothesis *could* mean that *any* principle which explains the conditioned reflex will also explain learning, it so far never *has* meant that. The CR hypothesis has always implied something akin to dominance as the explanatory principle. For that reason an examination of any other interpretation of the CR hypothesis is not strictly a legitimate function of this series.

In spite of the fact that our primary concern is with the CR hypothesis-as-explained-by-dominance, it seemed interesting and it was convenient to see if the temporal factor would account for other types of learning.

This additional inquiry provides the second line of approach mentioned at the outset of this paper. The results may be summarized as follows: When two lower-order reactions occur together, the stimulus presented first comes to take on the response of the second stimulus (**1**), (**2**) and (**5**). When, however, a lower-order and a higher-order reaction occur together the stimulus of the lower-order reaction takes on the response of the higher-order and the temporal factor seems to have no effect on this tendency (**3**). When two higher-order reactions are presented in close succession, as in the case of the study about to be reported, the second stimulus takes on the response of the stimulus presented first. This is, of course, directly opposite to what happened in the case of the two lower-order reactions.

Obviously the temporal factor which seems to determine

the direction of conditioning in the conditioned reflex proper, does not predict the outcome of other types of habit formation.

General Conclusions of the Series

From neither aspect of the examination does it appear that there is any known principle or combination of principles which will at one and the same time describe or predict the outcome of the conditioned reflex itself and of other types of habit formation. This stricture is, of course, in order only insofar as we have at one time or another used a valid criterion of dominance.

The Details of the Third Study: The General Problem

In this third study it seemed wise to use two higher-order reactions. In the first place we wished to see if dominance would predict the direction of conditioning when dominance was defined as importance, and this matter of importance seemed easier to control in the case of higher-order reactions. In explanation of our desire to measure the influence of importance we might repeat that in the second study of the series (**3**), the shock had come to elicit the reaction attached to the flash of light. Since the latter reaction was probably more important to the subject, in the sense that it quite likely involved most of his attention, it seemed that the CR hypothesis might stand if the concept of dominance were stretched to include importance.

In the second place two higher-order reactions seemed in line with the needs of the second aspect of our examination. The temporal factor had apparently determined the direction of conditioning in the case of the two lower-order reactions but not in the case of the combination of a lower-order and of a higher-order reaction. Having examined two lower-order reactions, and a combination of lower and higher, it was natural to see how the temporal factor could be applied to two higher-order reactions.

The Specific Problem

The essential part of the present test lay in arranging for two higher order reactions each with variable amounts of importance. To do this a target hitting task was assigned to each hand and provision made for weighting the successful hits of the respective hands so that (1) the contributions of the two reactions to the total score could be equal or that (2) one could be made to contribute more than the other.

The subject sat at a table on which there was a row of three metal contacts about an inch apart opposite his left hand and a similar row opposite his right hand. Normally the subject kept a stylus on the middle contact of the left row and another stylus on the middle contact of the right row. He was told to listen for the sound of a certain buzzer and then, with the left hand, to hit the other two contacts in a given order. If he hit quickly enough a bell with a distinctive tone would ring. Similarly, whenever another buzzer sounded he was to hit corresponding contacts with the stylus in his right hand. If he were successful with the right hand a bell with a different tone would ring. He was then told how much a success in the respective hands would count toward his total score. In the case of some subjects each bell (success) counted the same, while in the case of other subjects the bell for the left hand counted five times as much as the bell for the right, and in the case of still others the right counted five times as much as the left.

In addition to variations in the relative importance of the two tasks, the apparatus permitted variations in the order of presentation. Consequently it was possible to arrange that the more important task could be initiated before the less important (experimental group 1) or at the same time as the less important (experimental group 2) or after the less important (group 3); or that two tasks of the same importance could be initiated simultaneously (group 5) or that one of these equally important tasks could be initiated before the other (group 4).

Description of Apparatus.—The apparatus for producing those conditions is, in the essentials, the same as that described in detail in the first article of the series (2).

It was modified in the following ways: (*a*) Buzzers were substituted for the electrodes and a low voltage current used. The current for this buzzer passed through the stylus and middle contact to insure that the buzzer of either hand would not sound unless the stylus were on the middle contact. (*b*) A system of relays was introduced to make sure that the bell indicating the subject's success would not ring unless he hit the contacts in a given order and within a given time after the signal had sounded, and to make sure, further, that he could not make the 'success bell' ring again until after a second signal had been given.

Subjects.—Boys from grade 9 in a city junior high school were used as subjects. They were from 14 to 16 years old.

Explanation of Terms.—In order to avoid the repetition of clumsy locutions certain arbitrary terms will be used. These follow:

1. A *presentation* may refer to a pair of signals in close succession or to a single signal.

2. A *trial* is a presentation in which only one signal occurs.

3. The *more important hand* is the hand corresponding to the gong which has been assigned the greatest value. If making the left hand gong ring counts 5 and the right hand gong counts 1, then the left hand is the *more important hand* and the right hand is the *less important hand.* The terms less important or more important may also be applied to the respective signals.

4. The criterion for *conditioning* is that the subject when receiving only one signal shall move both hands or at least the hand for which no signal has been received. *No conditioning* is indicated in a trial to which the subject responds by moving only the hand corresponding to the given signal. A trial may indicate one of several types of conditioning. We will list the following types:

5. *L-M Conditioning.*—This means that the less important signal has come to elicit the response of the more important signal.

6. *M-L conditioning* is obvious from the above.

7. *1st-2nd (or First-second) conditioning* indicates that the first signal of a pair has come to elicit the response to the second signal.

8. The meaning of *2nd-1st conditioning* is obvious from the above. 2nd-1st conditioning is synonymous with the more customary phrase 'backward conditioning.'

9. *Left-right conditioning* indicates that the signal for the left hand has come to call forth the response for the right hand.

Procedure.—After the experimenter had explained and demonstrated the tasks, the subject took his seat in a three sided cubicle, with his back to the open side. He was informed of the respective values of the left and right bells and allowed to practise with no signal for a minute or so. He was then given some practice presentations consisting of trials to one hand at a time. These trials enabled the experimenter to tell if any conditioning were present before the experiment began. The series of trials was continued until the subject made the gong ring, after which a series of trials to the other hand was begun.

As soon as he had made the bell ring once in each hand the subject was told that the practice series was at an end and that the attempts from now on would count. A boy who had finished the experiment was set to keep a record of the number of times the subject made each bell ring, and a series of predetermined presentations began.

An irregular series of these presentations consisted of trials. For one half of the subjects the arrangement of trials was as follows: 11L, 16R, 20R, 26L, 29R, 34L, 36L, 39R, 42L, 50L, 60R, 66R, 74L, 79R, 86R, 89L, 96R. This indicates that the 11th pre-

sentation was a trial consisting of a signal to the left hand alone and that the 16th presentation consisted of a trial to the right hand alone and so forth. All presentations not listed above consisted of pairs of signals. For the other half of the subjects the sequence of left hand and right hand trials was reversed, the 11th presentation being a trial of the right hand signal and the 16th being a trial of the left hand. After about 100 presentations any conditioning which had been established in either direction was extinguished. Each subject served for only one sitting of about 25 minutes and was used in only one experimental group.

In order to avoid any influence of dextrality the rôles of more important-less important (or first-second) for one half the subjects were the reverse of what they were for the other half. Suppose that in one group the more important signal precedes the less important. In the case of five subjects of this group the left hand would be more important and come first while in the case of the other five subjects the right hand would be more important and come first. The trials of each of these sub-groups of five would vary from subject to subject, some of the five having the 11L, 16R etc., order and the rest having the 11R, 16L etc., order.

Results

Each subject was given trials of two types. For instance the subjects in Group 2 were given 8 or 9 trials with the more important signal and 8 or 9 with the less important. The answers to our questions lie in the extents to which these

TABLE I

Giving for Each Subject the Percentage of Trials of Each Type Which Show Conditioning

Group	1		2		3		4		5	
Description of Group	M.Imp. Signal Given First		M.Imp. and L.Imp. Signals Given Together		M.Imp. Signal Given Second		Signals Equally Important but One Given First		Signals Equally Important Given Together	
Type of Signal Tried	More Imp. (1st)	Less Imp. (2nd)	More Imp.	Less Imp.	More Imp. (2nd)	Less Imp. (1st)	1st	2nd	A	B
Percentage of trials showing conditioning	37	44	0	22	12	0	56	100	25	22
	22	62	37	22	44	25	67	12	78	100
	62	89	57	50	100	56	0	43	12	22
	0	12	11	12	78	100	37	56	22	62
	33	100	50	11	62	22	12	11	87	100
	0	50	50	22	50	11	75	100	37	22
	71	67	71	25	22	14	14	12	56	62
	12	0	56	62	37	33	22	75	11	25
	22	75	22	25	67	50	12	44	56	25
	62	89	87	100	50	11	56	100	100	89
Average	32	59	44	35	52	32	35	55	48	53
S.D.	25	31	26	26	25	28	25	37	30	32

different types of trials showed conditioning. This information is given by subjects [2] in Table I. Here a pair of numbers indicates the percentage of the two types of trials respectively which showed conditioning. In the case of the first subject in Group 1, for instance, 37 percent of the trials of the more important (which was also the first) signal showed conditioning, and 44 percent of the less important (second) signal showed conditioning. No subject appears in more than one group.

It will be observed that there is a considerable variation from subject to subject within any group.

In addition to variation, the matter of consistence is interesting. Some subjects seem liable to conditioning of either sort, some to neither sort and some to one type but not to the other. In general, however, a subject who shows great susceptibility to one type is also susceptible to the other type. The correlation between percentage of one type of trial showing conditioning and percentage of the other type showing conditioning is 0.69. This correlation is based on only the forty subjects of the first four experimental groups. In group five there were not two types of signals. This correlation between percentage of trials showing conditioning in the two types of trials does not mean, of course, that the subject showed as much success in one as in the other. It merely means that the subject who ranked high in his group in one type also ranked high in the other type. But the whole group may have averaged higher in one type than in the other.

Discussion of Results

Group 5 was used merely as a sort of control to see if the direction of the conditioning could be attributed to chance or to factors other than dominance or time order. In Group 5 two signals of the same importance were given at the same time. Hence if any factors other than the two mentioned were operating their influence would have a chance to show itself here. To check the influence of such possible irrelevant

[2] A mimeographed copy of the trial by trial results for each subject will be sent to anyone desiring such results.

factors, the left hand of five subjects (chosen by taking every other subject) was arbitrarily assigned to an hypothetical condition A and the right hand to condition B. In the case of the other five subjects the roles of right and left hands were reversed. It will be seen from Table I that in spite of this purely arbitrary separation, B has somewhat more conditioning than A, 53 percent of B and 48 percent of A trials showing conditioning. This preponderance of 5 percent was produced by a purely arbitrary division. Hence any other preponderance of 5 percent in favor of one type of signal cannot be regarded as significant. We will speak of differences of this order as 'chance preponderances.' By this we do not imply of course that 5 percent represents the greatest difference that we could expect from chance. It is merely an empirical base line.

In the tables that follow, certain supplementary groups not heretofore mentioned are introduced. These are designated by letters. Group A supplements Group 1, Group B supplements Group 2 and so forth. These supplementary groups were arranged because it may be questioned that the task to which the greatest weight was attached was always the most important task. It may be that the subject did not concentrate his attention on this task. He may have accidentally achieved more successes with the less important task and may have regarded that task as easier. If that were true it is quite possible that he concentrated his efforts on the easier task, thinking that it is better to work for an easily obtained '1' than for a difficult '5.'

In order to check on that possibility we used the number of successes as the criterion of importance. If the subject achieved more successes with the left hand it was considered more important in his case regardless of the weight attached to a success in that hand. It will be noticed that the number of cases in the various groups was altered considerably, Group D (and possibly C) having shrunk so much that its value is questionable.

Unless otherwise noticed, the discussion will refer to the groups formed by the original criterion. Discussion of the

supplementary groups will be reserved until later. Data referring to supplementary groups are given in parentheses.

In Table II the data are arranged so as to show the extent to which the L–M conditioning expected by the CR hypothesis has taken place.

TABLE II[3]

Description of Groups	Designation of Groups		No. of Cases	Percentage of Trials Showing L–M Conditioning	Percentage of Trials Showing M–L Conditioning	Difference in Favor of L–M Conditioning
	Orig. Criter.	Supp. Criter.				
More important signal coming first	1	(A)	10 (19)	58 (55)	31 (28)	27 (27)
Both signals coming together	2	(B)	10 (18)	36 (51)	43 (43)	− 7 (8)
More important signal coming second	3	(C)	10 (6)	32 (50)	52 (67)	−20 (−17)

Table II shows rather clearly that L–M conditioning is not the rule. We get a preponderance of this type of conditioning when the more important signal comes first, mere chance preponderance when more important and less important signals come together and a preponderance of the other type when the more important signals succeeds the less important. It seems as if some other factor were responsible for whatever preponderance there is.

The other factor which varied systematically from group to group was the time interval. Table III shows the relation between this temporal factor and any preponderance that there may be.

In Table III we notice (with reference to Groups 1, 4 and 3), a consistent preponderance of 2nd-1st (backward) conditioning. There is much more likelihood of the second stimulus calling forth the response of the first than there is of

[3] The figures given in Tables II and III may differ somewhat from the averages given in Table I. The values given in Tables II and III were determined from the percentage of the total trials. Those in Table I from the average of the individual percentages.

the first calling forth the response of the second. This holds true no matter how the first and second signals may differ in importance.

TABLE III

Description of Groups	Orig. Criter.	Supp. Criter.	No. of Cases	Percentage of Trials Showing 2nd–1st (Backward) Conditioning	Percentage of Trials Showing 1st–2nd Conditioning	Difference in Favor of 2nd–1st Conditioning
Second signal less important	1	(A)	10 (19)	58 (55)	31 (28)	27 (27)
Both signals of same importance	4	(D)	10 (3)	55 (52)	37 (46)	18 (6)
Second signal more important	3	(C)	10 (6)	52 (67)	32 (50)	20 (17)

Such a marked tendency toward 2nd-1st conditioning is rather surprising. While 2nd-1st or backward conditioning has been reported in other cases it was at one time thought impossible and is even yet considered *rara avis*. Here, however, it is not only observed but it is the rule—the only rule we could depend on in predicting the predominant direction of conditioning.

The relative rôles of importance and of time order may be clearer from a numerical statement of the expectations to be inferred from the results of this experiment. Suppose that stimuli A and B are presented in conjunction with one another and that we wish to know to what extent A will come to take on the response to B. If A is habitually presented after B it will come to take on the response to B in from 50 percent to 60 percent of the trials. If it is presented before or at the same time as B it will take on B's response in 30 percent to 40 percent of the trials. If in addition to succeeding B in time, A is less important than B, its likelihood of taking on the response to B may be increased from about 55 percent to 58 percent. If A is more important than B that likelihood may be decreased from 55 percent to 52 percent. These figures show a tendency to increase slightly as the first signal becomes more important. We should hasten to add that

this consistency is the only justification for mentioning such small differences since they are less than chance preponderance. Even two such differences in succession must be regarded as merely suggestive.

From the above it is rather clear that in this study the most important factor in determining the direction of the conditioning is the time order, and that the influence of this factor is directly opposite to what it has been in previous studies.

The results yielded by the second criterion are very similar to those of the first. In both cases we get a regular preponderance of 2nd-1st conditioning and no such preponderance of L–M conditioning.

The differences between the results of the two criteria are of minor importance. The difference between the negative preponderance of L–M conditioning in Group 2 and the positive preponderance in the corresponding Group B becomes less significant when we remember that both these preponderances are in the neighborhood of chance. I would attach greater weight to the positive preponderance of Group B on account of the greater number of cases in that group. The discrepancy in Table III between the results of Group 4 and the corresponding Group D is likely due to the small number of cases in Group D.

General Conclusions of the Third Study

When two higher-order reactions occur repeatedly in conjunction with one another, the following results appear:

1. Both types of conditioning will appear to some extent. Each of the two stimuli will come to elicit in some cases the response originally attached to the other stimulus.

2. Although both types will appear, the type described as second-first, or backward, conditioning predominates.

3. The relative importance of the two reactions (when determined by two distinct criteria of 'importance') has very little, if any, influence.

(Manuscript received October 13, 1934)

References

1. Bernstein, A. L., Temporal factors in the formation of conditioned eyelid reactions in human subjects, *J. General Psychol.*, 1934, **10**, 173–197.
2. Stephens, J. M., The conditioned reflex as the explanation of habit formation: I. The essential factors in the establishment of the conditioned reflex, *J. General Psychol.*, 1934, **10**, 110–136.
3. Stephens, J. M., The conditioned reflex as the explanation of habit formation: II. The operation of a higher-order reaction and a lower-order reaction in close succession, *J. General Psychol.*, 1934, **10**, 219–227.
4. Razran, H. S., Theory of conditioning and of related phenomena, *Psychol. Rev.*, 1930, **37**, 25–43.
5. Wendt, G. R., Analytical study of the conditioned knee jerk, *Arch. Psychol.*, 1930, **18**, No. 123, pp. 97.

CHANGES IN NEURO-MUSCULAR TENSION ACCOMPANYING THE PERFORMANCE OF A LEARNING PROBLEM INVOLVING CONSTANT CHOICE TIME [1]

BY EDWIN GHISELLI
University of California

Previous investigators (1) (3) concerned with studying the changes in neuro-muscular tension occurring during learning have employed tasks in which unlimited choice or response time was allowed the subject. These tasks may be referred to as 'power tasks.' The present problem is concerned with the measurement of the change in neuro-muscular tension of the active hand during the process of performing a 'speed task.' The task selected for learning was a serial multiple-choice problem involving a short eye-hand coördination. The time interval allowed for making each response in the task was constant so that the total possible time of reacting was the same for all subjects. This made the task a 'speed task' as contrasted with those tasks wherein the subject's response is not restricted to a definite time interval.

It was proposed to investigate the changes in absolute pressure and the changes in the variability of pressure under the following conditions: 1, changes in tension during learning; 2, changes in tension during over learning; and 3, changes in tension during the learning of a second similar task immediately following the learning of an initial task.

Apparatus and Procedure.—Two serial multiple-choice problems, involving the same kind of eye-hand coördination, were presented to each subject. The task required of the subject was to react to the flash of a small signal light by depressing one of three plunger keys. The keys operated vertically and the pressure was always applied with the right forefinger. The light flashed on and off in a series of seven flashes, each series constituting a trial. The duration of each flash was one second with a two second interval between flashes. The time between trials was six seconds. For each flash of the signal light the depression of only one of the keys constituted a

[1] The writer wishes to express his indebtedness to Dr. C. W. Brown under whose direction this experiment was performed.

correct response. Pressure upon either of the other keys constituted an error and was indicated by the appearance of a flash from another light directly below the signal light. The correct key was varied among the three keys during a trial and only one response was allowed for any one flash of the signal light.

Through a rigid system of steel levers and springs the extent to which the keys were depressed was recorded on a polygraph. A linear relation was observed between the kilograms-pressure applied to the keys and the extent of movement of the recording pen. Thus, the extent of movement with which the keys were depressed could be translated directly into kilograms-pressure. The keys could be depressed to the extent of six millimeters. However, in no instance during learning did a subject push down any key the full distance.

Two sequences of correct keys were used. In problem A the sequence of correct keys was right, left, center, left, right, center, left; and for problem B, left, center, right, center, left, right, center. Each subject was given twenty trials on one of the sequences of correct keys, and following a one minute rest, twenty trials on the other sequence. A counterbalanced order of presenting the problems was used, one group of subjects having problem A first and problem B second, another group having problem B first and problem A second. The subjects were twenty-two graduate and undergraduate students, none of whom had had previous experience with such a problem. Each group consisted of four women and seven men.

Results.—Inasmuch as the subjects reached the criterion of learning, three consecutive errorless trials, at different times, it was necessary to reduce their learning times to a common basis in order to determine the amounts of pressure occurring at comparable stages in learning. This was done by the Vincent method. Graphs were made for each of the following relationships for each of the subjects: the number of correct choices per trial, the average pressure per trial, and the standard deviation of the pressure per trial. Each curve was smoothed by the graphical method and the learning period divided into ten equal parts. The value at the midpoint in each of the dependent variables was read from the graphs of every decile division. Since the number of choices for each trial was only seven, Fisher's (2) formula for the estimated standard deviation when the number of cases is small [2] was used.

The subjects who learned problem A first required on an average 9.6 trials to learn this problem, and 9.4 trials to learn problem B. The other group of subjects required 9.9 trials to learn problem B, and 9.4 trials to learn problem A. Since the problems were of equal difficulty, regardless of the order

[2] $s^2 = \frac{S(x - \bar{x})^2}{n - 1}$, Fisher, R. *Statistical methods for research workers.* Pg. 46.

in which they were presented, the results obtained from both groups on the first problem learned were combined. Similarly, the results from both groups on the second problem learned were combined.

1. *Changes in Pressure During Learning.*

a. Changes in Absolute Pressure.

Bills (1) and Stroud (3) in their studies of stylus maze learning report that there is a gradual decline in pressure on the stylus as learning proceeds. In the present experiment the Vincent values of absolute pressure on the keys indicate that the pressure tends to be maintained at a relatively constant level throughout the entire period of learning (Table I). Figure 1 gives a graphic representation of the average pres-

TABLE I

VINCENT VALUES OF THE MEAN OF THE ABSOLUTE PRESSURE EXPRESSED IN KILOGRAMS

Deciles	1	2	3	4	5	6	7	8	9	10
First Problem	2.18	2.19	2.12	2.14	2.14	2.14	2.10	2.19	2.07	2.06
Second Problem	2.06	2.10	2.10	2.12	2.12	2.11	2.08	2.08	2.08	2.10

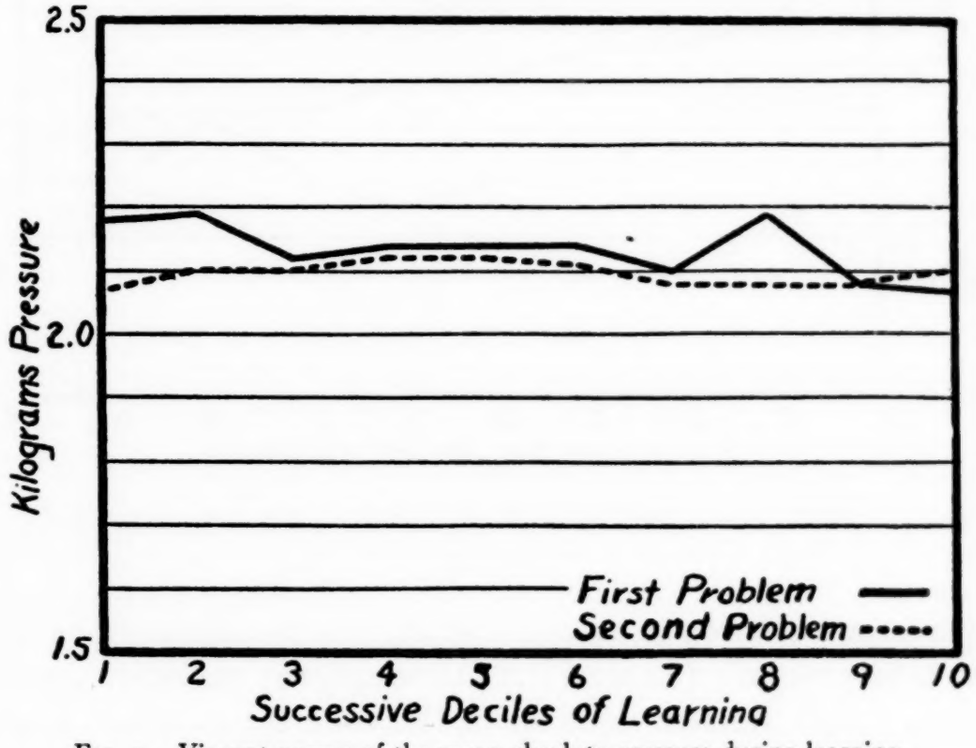

FIG. 1. Vincent curves of the mean absolute pressure during learning.

sure of the 22 subjects for each problem. No consistent change is noticeable in either curve. An examination of each individual's curve of learning, for each of the problems, shows the same general tendency, no consistent trends occurring in the record of any subject. It would seem then, that under the conditions of this experiment in which the time interval for each choice response is held constant, there is no decrement in tension such as characterizes tasks wherein unlimited choice time is allowed.

b. Changes in Variability of Pressure.

Variability of tension during learning has received no notice by previous investigators. The variability of pressure in the first problem undertaken, represented by the solid line in Fig. 2, shows that the initiation of this task is characterized

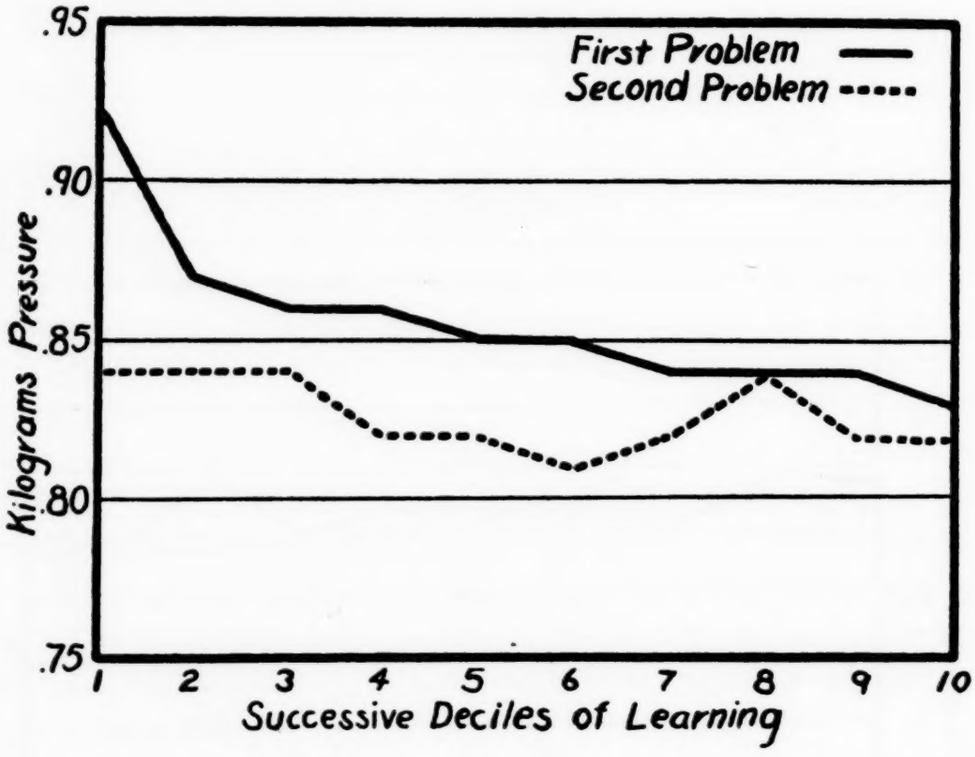

FIG. 2. Vincent curves of the mean of the individual standard deviations during learning.

by a large variation in pressure which rapidly decreases as the learning proceeds (Table II). Although the variability continues to get smaller right up to the time the criterion is at-

TABLE II

VINCENT VALUES OF THE MEAN OF THE INDIVIDUAL STANDARD DEVIATIONS OF PRESSURE EXPRESSED IN KILOGRAMS

Deciles	1	2	3	4	5	6	7	8	9	10
First Problem	.92	.81	.86	.86	.85	.85	.84	.84	.84	.83
Second Problem	.84	.84	.84	.82	.82	.81	.82	.84	.82	.82

tained, two thirds of the decrease is accomplished in the first third of the learning period.

2. *Changes in Pressure During Overlearning.*

a. Changes in Absolute Pressure.

In order to include all the subjects in this analysis only the first four trials of overlearning were used. When these are considered, together with the three errorless trials of the criterion, there are seven errorless trials in which changes in pressure are noted. The results indicate a marked decrement in the absolute pressure during the overlearning of each problem (Table III). A clearer picture can be obtained from the curve of Fig. 3. The rate of decrement is very marked, es-

TABLE III

MEAN OF THE ABSOLUTE PRESSURE EXPRESSED IN KILOGRAMS PER TRIAL DURING THE CRITERION OF LEARNING AND OVERLEARNING

Trials	1	2	3	4	5	6	7
First Problem	2.06	1.99	1.91	1.87	1.83	1.76	1.72
Second Problem	2.10	1.89	1.86	1.76	1.75	1.79	1.74

pecially when it is compared with the relatively constant level maintained during learning. A study of the overlearning of those subjects who had more than four trials after satisfying the criterion, indicates that the decrement continues right up to the last trial of this prolonged overlearning period. The rate of decrement, of course, becomes smaller and smaller as overlearning proceeds.

b. Changes in Variability of Pressure.

From Table IV and Fig. 4 it will be seen that during the period of overlearning the variability of pressure is consis-

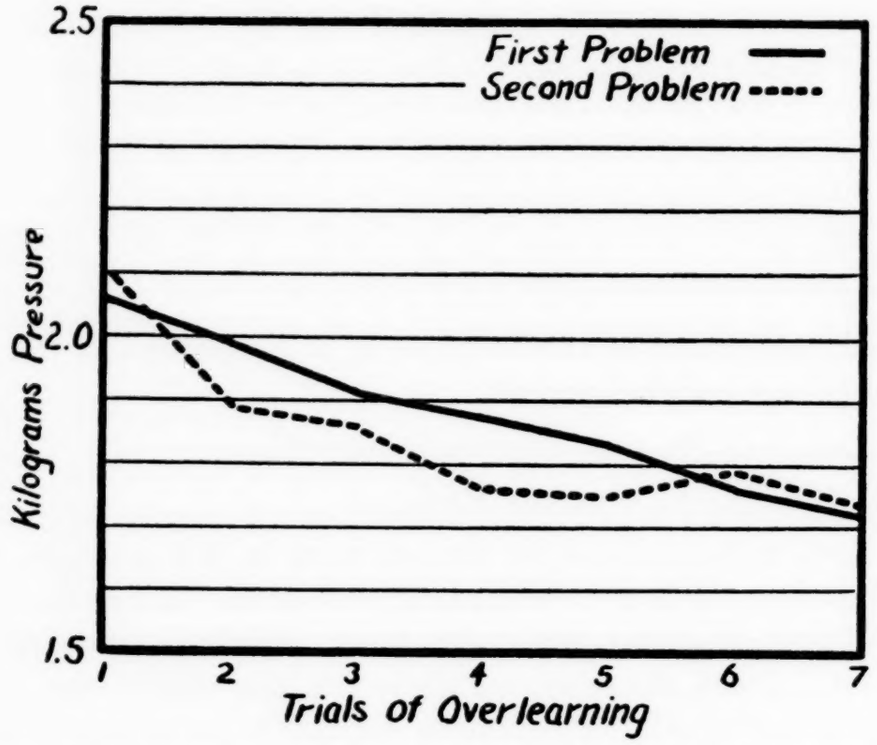

FIG. 3. Mean absolute pressure during the criterion trials and the first four trials of overlearning.

TABLE IV

MEAN OF THE INDIVIDUAL STANDARD DEVIATIONS OF PRESSURE EXPRESSED IN KILOGRAMS PER TRIAL DURING THE CRITERION OF LEARNING AND OVERLEARNING

Trials	1	2	3	4	5	6	7
First Problem	.80	.77	.80	.75	.76	.75	.80
Second Problem	.77	.77	.77	.75	.80	.79	.75

tently small. This is characteristic of both the first and second problems learned.

3. *Changes in Pressure During the Learning of a Second Task as Compared with the Change in Pressure in a Similar Task just Preceding it.*

a. Changes in Absolute Pressure.

A graphic representation of the tension occurring in the second problem learned as compared with the first problem is found in Fig. 1. It will be seen that the tension during the second problem remains relatively constant at a level of pres-

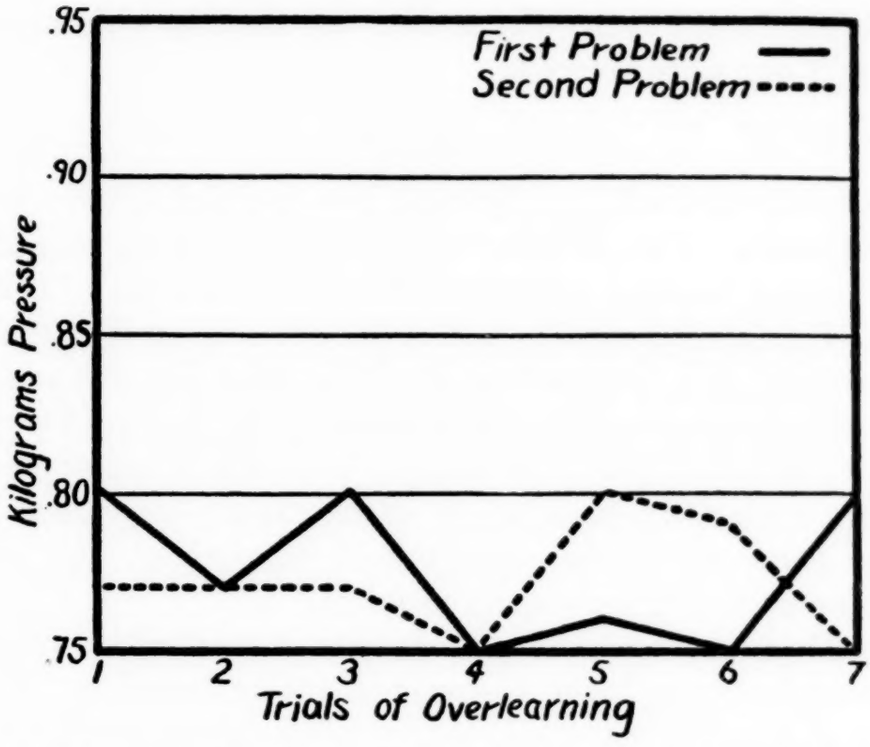

FIG. 4. Mean of the individual standard deviations during the criterion trials and the first four trials of overlearning.

sure approximately the same as for the first problem, the minor fluctuations being even smaller than those found in the first problem. Although previous investigations have not included situations comparable to the present one, it might be expected that the tension manifested in a new problem would begin at a high level and gradually decrease as the learning continued. In the present experiment the tension occurring in the initial adaptation to the general experimental situation demaded by the first problem or in the initial adaptation to the specific task of the second problem is no greater than the tension manifest throughout the remaining stages of learning.

b. Changes in Variability of Pressure.

A comparison of the curves of variability of pressure for the two problems, Fig. 2, shows that the large variations of pressure which accompanied the initial reactions to the first problem do not occur in the second problem. The level of variability of the second problem remains approximately at the level attained at the end of the first problem. There is a

tendency for the variability to decrease slightly with learning, but it is not very marked.

Summary

Twenty-two subjects learned two multiple-choice visuo-motor tasks. The pressure on the response keys was measured during learning and overlearning. It was found that:

1. During the learning period the pressure on the keys was maintained at a constant level. The level was the same in second problem learned as was found in the first problem.

2. The initiation of the first problem was accompanied by a large variability of pressure which rapidly decreased to a lower level which was maintained throughout the whole learning period of the second problem.

3. During overlearning there was a marked decrease in pressure from the level maintained during learning. Variability of pressure was consistently small during overlearning.

Conclusions

In the light of the above results it appears justifiable to conclude that neuromuscular tension is maintained at a constant level during the learning of a short speed task which demands constant application on the part of the learner. Once the task is learned, however, the tension markedly decreases. Adaptation to such a task is manifest by a rapid decrease in variability of tension during learning, and a consistent small variability during the subsequent learning of a similar problem.

(Manuscript received November 19, 1934)

Bibliography

1. Bills, A., Tensions in learning and association, *Proc. Ninth Int. Cong. Psychol.*, 1929, 75–76.
2. Fisher, R., *Statistical methods for research workers.* London. 1932.
3. Stroud, J., The role of muscle tensions in stylus maze learning, *J. Exper. Psychol.*, 1931, **14**, 606–631.

THE EFFECT OF ART TRAINING ON MIRROR DRAWING

BY CHARLES D. FLORY
University of Chicago

A large share of human learning consists of activities incident to the acquisition of skills. The processes involved in becoming skilled in any particular act have been designated as sensori-motor learning. Many experiments have been performed to determine the difficulties encountered by the learner in mastering the various kinds of skills. Some variation of standard mirror drawing experiments has been a favorite sensori-motor exercise for classes in educational psychology. It is the purpose of this paper to consider the effect of art training in a particular sensori-motor situation. Are students who have had training in art able to learn to trace a mirrored image more quickly than students who have had no art training?

Procedure

Twenty-six students in an educational psychology class were used as subjects in the experiment. A six-pointed star in outline, 2⅞ inches from opposite points, was the basis of the learning exercise. Each student was provided with thirty-one symmetrical stars. The subjects were instructed to trace thirty of these stars while viewing the image through a mirror. A preliminary trial was made without the mirror to determine the time required for the task under normal conditions. The time spent on this first tracing was considered the normal speed or control time for each subject. All tracings were made in a clockwise direction.

Thirty trials were made by each subject under mirror drawing conditions. All tracings began at the point nearest the subject's body. Since the experiment was planned merely to illustrate some of the problems encountered in sensori-motor learning, each student was allowed to arrange his own equipment and do the work at home with the help of an assistant. The assistant acted as timekeeper and recorder for the subject. Each tracing was considered a trial since there were no practice periods between tracings.

The class was divided into three groups for the experiment. Group I did the entire exercise at a single sitting. Group II did three tracings at each sitting with a twenty-four hour interval between sittings. Group III spaced the trials eight hours apart over a period of ten days. These three methods of procedure were used to demonstrate the effect of concentrated and divided practice periods.

Seven students in the class were also enrolled at the Chicago Art Institute while the experiment was in progress. Nine subjects had received some training in art above the elementary school, while eleven members of the class reported no training in art. The seven students from the Chicago Art Institute were matched with seven subjects who had no art training above the elementary school. Since three methods of procedure were used in the experiment it was necessary to pair each art student with a classmate who had used the same technique. Sex differences in drawing skill were also controlled by paring the subjects. Three of the art students did the experiment at one sitting with two each using the second and third procedures. This group of seven art students will be referred to as the trained group and the seven students with whom they were paired will be known as the untrained group. It was impossible to make the groups exactly comparable since the problem arose after the exercise had been completed. There are reasons however for believing that most of the external factors have been controlled so that art training is the significant variable between the two groups.

Results

The average time required for the controlled tracings was calculated for the three types of subjects used in the experiment. It seems significant that the time required for these original tracings without the mirror was 25, 26, and 29 seconds [1] respectively for art students, students with some art training, and students with no art training. Tracing skill appears to be related to the amount of training in art. But the real problem under consideration is the effect of art training on mirror drawing. The remainder of the discussion will consider the learning of the trained and untrained groups.

The average time required for these two groups to trace the stars on each trial is given in Table I. Absolute and smoothed means have been presented. An inspection of these data indicates that the two groups were vastly different in mirror drawing skill at the start. The data reveal that the two groups were nearly equal in tracing ability at the end of the experiment. Differences in the learning of the two groups are shown quite strikingly by the curves in Fig. 1.

It is indeed significant that the trained group was nearly twice as efficient in mirror drawing as the untrained group on the first trial. This fact becomes more significant when it is recalled that there was an average difference of only four

[1] Miss Bernice Weinhoff, a member of the class, made the tabulations and calculations from which the present interpretations have been suggested. All averages reported have been rounded to the nearest second.

seconds between the groups in their controlled tracings. Art students seem to have mastered a technique which enabled them to adjust more quickly to this new drawing situation.

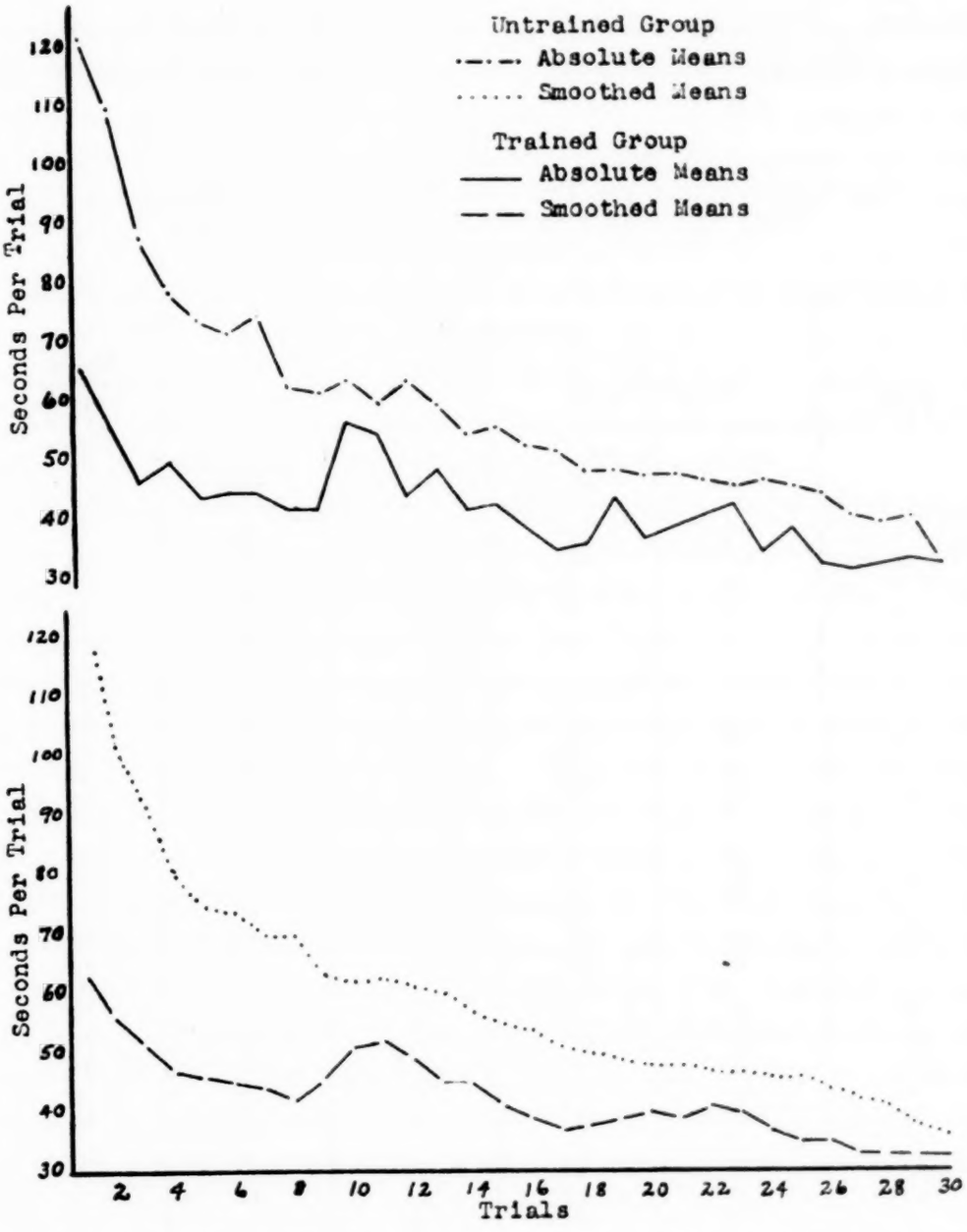

FIG. 1. Learning curves for two groups of students in a mirror drawing experiment to show the effect of art training on skill in mirror drawing.

The untrained group reduced its tracing time from 122 to 63 seconds in eight trials, an improvement of 48 percent. The trained group started with a degree of efficiency

which the untrained group reached only after eight attempts. In an equal number of trials the trained group had diminished its time from 61 to 41 seconds or to a point within 16 seconds of the average time required by this group on the controlled tracing. Thus the trained group reduced its time by 38 per-cent while the untrained group was reducing its time 48

TABLE I

THE TIME REQUIRED PER TRIAL FOR TWO GROUPS OF STUDENTS TO TRACE AN OUTLINE SIX-POINTED STAR AS REFLECTED IN A MIRROR

Trials	Seconds per Trial				Absolute Difference Between the Two Groups
	Untrained Group		Trained Group		
	Absolute Mean	Smoothed Mean	Absolute Mean	Smoothed Mean	
1	122	118	66	63	56
2	110	103	56	56	54
3	88	92	47	51	41
4	79	80	50	47	29
5	74	75	44	46	30
6	72	74	45	45	27
7	75	70	45	44	20
8	63	70	41	42	22
9	62	63	41	46	21
10	64	62	57	51	7
11	60	63	55	52	5
12	64	61	44	49	20
13	66	60	49	45	11
14	55	57	42	45	13
15	56	55	43	41	13
16	53	54	39	39	14
17	52	51	35	37	17
18	49	50	36	38	13
19	49	49	44	39	5
20	48	48	37	40	12
21	48	48	39	39	9
22	47	47	41	41	6
23	46	47	43	40	3
24	47	46	35	37	12
25	46	46	39	35	7
26	45	44	33	35	12
27	41	42	32	33	9
28	40	41	33	33	7
29	41	38	34	33	7
30	33	36	33	33	0

percent. After eight trials the trained group was still 22 seconds more efficient than the untrained group.

From trial nine to the end of the experiment the two groups were more alike in their performance than during the early part of the exercise. In fact on trial eleven the trained group was only 5 seconds faster than the untrained group, which means that the tracing time for the two groups was within one second of the difference between them on the controlled trial. This similarity in performance on trial eleven was due both to an increased efficiency for the untrained group and a decreased efficiency for the trained group. There were more fluctuations in the average tracing time of art students than for students who had no training in art. The fluctuations suggest that the task presented different problems to the two groups of subjects.

A lapse in the efficiency of the trained group on trial ten and again from the nineteenth to the twenty-third trials inclusive suggests several possible conditioning factors. The so-called 'artistic temperament' has been thought by some psychologists to produce an erratic type of performance. It is possible that art students began the experiment with more zeal than the untrained group. This additional enthusiasm may have produced better results in the early trials than would have been expected under normal conditions. An artist has suggested that the novelty of the task would have been exhausted for artistic people by the time eight or nine attempts had been made. Progress by the trained group after trial nine may have been the result of forced activity as contrasted with spontaneous practice from trial one to nine inclusive. Boredom could easily have produced an erratic type of performance such as the learning curve for art students reveals. Verbal reports from the trained group suggest interest rather than boredom throughout the experiment. It may be that art students were consciously trying to make a good showing on a task which was closely akin to their usual work. Overanxiety to do well may have produced an emotional state which resulted in an irregular type of learning curve. Artists may have attempted to analyze the problem

into its spatial elements and thereby encountered a distraction which produced temporary lapses in efficiency during the course of the learning.

The learning curve for the trained group flattens out from trial twenty-six to the end of the exercise. It may be that 33 seconds is approximately the physiological limit for performing the task set in the experiment. The fact that the trained group was only 8 seconds slower with the mirror than on their controlled tracing lends support to this hypothesis. As the efficiency of the group approached the physiological limit one would expect some irregularities in the learning curve. It is a well known fact that skilled artists are not up to their best performance on every trial in any ordinary motor task. The artistic temperament, the artistic possibilities of the task, attempted analysis of the spatial elements involved, and the nearness of the performance to the physiological limit are factors which should be investigated with a larger number of subjects under rigidly controlled conditions to determine the effect of each factor on a sensori-motor skill such as mirror drawing.

The trained group seems to have mastered the difficulties presented by the mirror. The final efficiency of art students did not equal the performance under normal conditions but one would hardly expect mirror tracing time to equal ordinary tracing. It is interesting to note that the untrained group did the last trial in 33 seconds. This sharp drop on the last trial suggests a psychological effect of the knowledge that no more chances would be allowed in which the subjects could make a good showing. It is clear that the untrained group needed more than thirty trials to bring their learning to a plateau similar to that reached by the art students.

Tentative Conclusions

It is dangerous to draw conclusions from data which are an inadequate sample of the classes considered. The present report is however only suggestive and the conclusions are to be considered as tentative.

1. It appears that art training produces a type of ability or skill which transfers to a learning situation involving a mirrored image.

2. Students trained in art were nearly twice as efficient in mirror drawing at the start as the untrained group.

3. Both groups increased rapidly in efficiency during the early trials. Thus skill in one task does not mean that practice in related skills is unnecessary.

4. It is possible that art students were bored more than untrained students after the artistic possibilities of the design had been exhausted. The nearness of the practice time to the control time for the trained group may have produced the erratic learning curve during the last two-thirds of the experiment. An attempt on the part of art students to analyze the spatial elements involved may have produced temporary lapses in efficiency.

5. Untrained students can reach the efficiency of the trained group if sufficient time is allowed and sufficient effort is expended.

These results suggest several fruitful problems in connection with mirror drawing which should be investigated. The reasons for the fluctuations in the learning of art students should receive attention. Considerable transfer seems to have resulted from training in art. It would be of value to determine the types of abilities or skills which are facilitated by art training. The problem should be studied with more subjects and under rigorously controlled conditions.

(Manuscript received November 22, 1934)

CERTAIN SOMATIC ACTIVITIES IN RELATION TO SUCCESSFUL AND UNSUCCESSFUL PROBLEM SOLVING

BY MYRON S. CLITES

Indiana University

PART II

SKIN RESISTANCE TO AN ELECTRIC CURRENT IN RELATION TO SUCCESSFUL AND UNSUCCESSFUL PROBLEM SOLVING

Electrical resistance taken on the palmar surfaces of the hand correlates very closely with activity of the sweat glands, and serves, therefore, as an indicator of sympathetic activity. In this study an attempt is made to determine whether sympathetic activity is increased or decreased during successful and unsuccessful problem solving. Darrow [1] reports that sensory stimulation is accompanied by larger drops in electrical resistance than is ideational stimulation. In spite of the smaller drop in resistance accompanying ideational stimulation, it would be interesting to determine whether or not there is a lower or higher resistance level maintained during successful as compared with unsuccessful problem solving. Davis [2] and Collman [4] report that superior people have a consistently lower resistance than do mental deficients. Sears [5] found that an abrupt increase in complexity or speed of addition, was accompanied by a large increase in amplitude of deflection, as compared with the immediately preceding examples; that abrupt decreases in complexity or speed was accompanied by a decrease in amplitude of deflection, though these decreases are in no case great; and, that despite the adjustment curve tendency, deflections for all the difficult examples and for the majority of fast examples, regardless of position in the series, are larger than the mean deflection for the easy or slow examples. Hence, although sudden appearance of a more or less critical situation seems to be the main

factor in causing large deflections in his addition experiments, a more continuous increase in the energy or tension-evoking character of the stimulus situation appears to be a contributing factor. This latter statement would lead one to suspect that there would be a difference in the resistance levels when the presentation of a problem is responded to by successful solution, giving up, or a failing attempt. It is our purpose to determine whether or not such a phenomenon occurs.

The problem used was the second water dipping problem chosen from the adult level of the Stanford Revision of the Binet Test and the apparatus used for measuring the skin resistance was one described by Davis and Porter [3]. The subject was seated behind a screen where he could not see the recording apparatus or the experimenter. S placed his forefinger and his middle finger so as to cover the nails in separate clay cups filled with physiological salt solution. The clay cups were in a zinc sulphate electrolyte and zinc plates were dipped into this electrolyte.

During one sitting, E took resistance readings at the expiration of each minute for five minutes of rest. Then E told S to read the directions and problem from a sheet of paper which had previously been placed in his hand. S told E when he had finished reading problem and was ready to start solving. E took resistance reading just as soon as needles became steady after signal to read had been given. Then E caught the reading just previous to S's informing E that S had finished reading problem. During solving, E took readings each minute and caught reading just before S informed E that he had solved the problem or was ready to give up. After approximately five minutes, if S had made no indication, E told S not to bother to work on the problem any more and assured him that the purpose was merely to get some records during mental rest and mental effort. In case of a solution, E took a reading just after S started reporting and just prior to close of reporting. After report (or giving-up) there was another rest period of five minutes during which E took readings each minute.

The situation in the case of the 'control' sitting was just as in the problem solving situation except that there was no problem to be read, or to be solved, and no report of solutions to be made.

TABLE I

DIFFERENCE IN RESISTANCE IN THE 'CONTROL' AND PROBLEM SITUATIONS AT THE POINTS INDICATED. RECORD FOR SUCCESSFUL SUBJECTS
Resistance is higher during the 'control' situation.

	Difference between 'Control' and Problem				
	Beginning		Middle		End
Rest	2626	2359	1877	2536	3250
Reading	3747				4092
Solving	4092		4127		4395
Reporting	4923				4864
Final rest	4178	4208	5022	5061	3993
	Correlation of 'Control' and Problem				
	Beginning		Middle		End
Rest	+.64	+.62	+.65	+.61	+.68
Reading	+.59				+.41
Solving	+.41		+.49		+.44
Reporting	+.37				+.29
Final rest	+.43	+.53	+.41	+.46	+.66
	S.D. of Each of the r's				
	Beginning		Middle		End
Rest	±.141	±.147	±.138	±.149	±.128
Reading	±.155				±.198
Solving	±.153		±.181		±.192
Reporting	±.206				±.218
Final rest	±.194	±.174	±.203	±.188	±.134
	$\frac{\text{Diff.}}{\text{S.D. (Diff.)}}$ Between 'Control' and Problem				
	Beginning		Middle		End
Rest	1.9	1.9	1.6	2.0	2.5
Reading	3.4				3.1
Solving	3.1		4.6		3.3
Reporting	3.4				2.9
Final rest	3.3	2.7	3.5	3.6	4.2

The 'control' sitting was given first for half the subjects while the procedure was reversed for the other subjects.

The ammeter and voltmeter readings were used as bases for computing the resistance in ohms, calibration constants of the instrument being empirically checked from time to time.

Average resistance for the groups of subjects, were computed by minutes where possible. For the solving period it was impossible to adhere to a strict time basis in grouping results, since not all subjects worked the same length of time. Therefore resistance at the beginning, in the middle, and at the end of solving period was considered comparable for the various subjects.

TABLE II

DIFFERENCE IN RESISTANCE IN THE 'CONTROL' AND PROBLEM SITUATIONS AT THE POINTS INDICATED. RECORD FOR FAILING SUBJECTS

Resistance is higher during the problem situation except during attempted solution.

	Difference Between 'Control' and Problem Situation				
	Beginning		Middle		End
Initial rest.....	2357	2273	2247	2935	2958
End of rest to reading in problem situation.....		4271			
	Correlation of 'Control' and Problem Situations				
	Beginning		Middle		End
Initial rest.....	+.110	+.050	+.210	+.210	+.160
$S.D._r$	±.197	±.222	±.212	±.212	±.217
	$\frac{\text{Diff.}}{\text{S.D. (Diff.)}}$ of 'Control' and Problem Situations During Initial Rest				
	Beginning		Middle		End
Initial rest.....	1.2	1.4	1.0	1.5	1.5

In the case of successful subjects the resistance in the 'control' procedure is continuously above that during the problem solving situation. This is true during initial rest, reading of problem, solving, reporting, and final rest. The

average of the correlations at the different stations is + .52, so any group tendency is also a tendency of the individuals composing the group. Inspection of the data reveals that a large majority of the subjects maintained higher resistance during the control procedure than during problem solving. The resistances for those few whose readings were the reverse of the general trends were about the same during the two procedures.

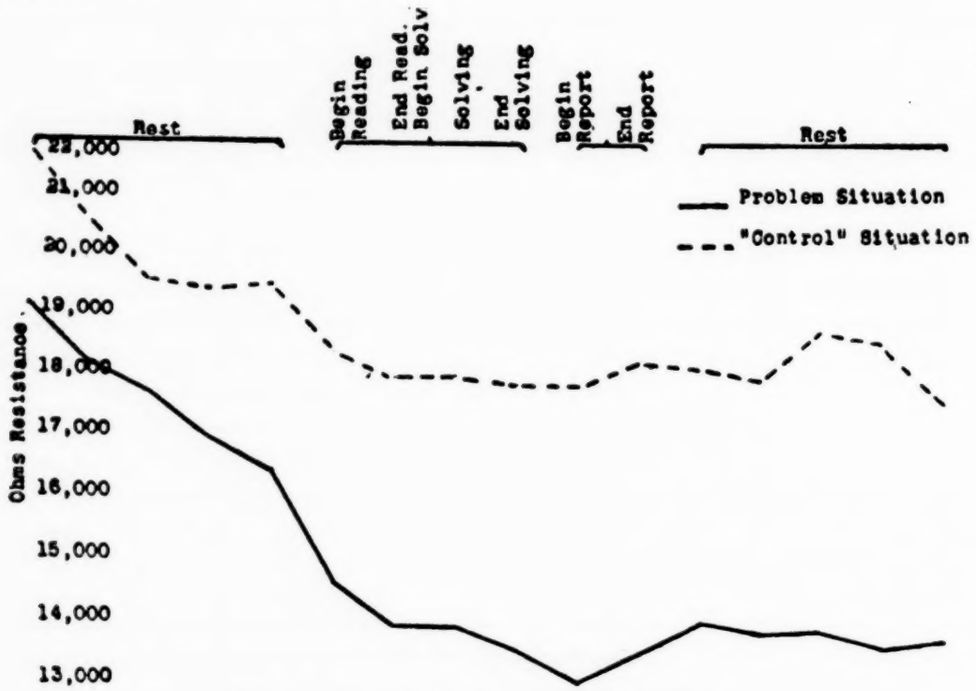

FIG. 1. Record for successes (18 subjects).

During the period of initial rest the control readings are higher than the problem readings for successful subjects while the opposite is true of the unsuccessful subjects. When the two differences which are opposite in direction are added and divided by the S. D. (Diff.) the quotients range from 2.6 to 3.5 and indicate that the chances are very remote that the resistance readings for successful and failing subjects will ever become reversed.

As soon as the successful subjects start to read the problem the resistance drops so that it is 3747 ohms lower than for the corresponding control situation. The difference at this point is reliable and this reliability continues throughout the sitting.

During the reading of the problem and attempted solution by the failing people the resistances are practically the same as in the corresponding control situation. During the final rest the resistances diverge slightly and take up a position similar to that during the first rest period but the resistance readings during the final rest differ very slightly.

When the failing subjects are told to start reading the problem there is an average drop of 4271 ohms in resistance.

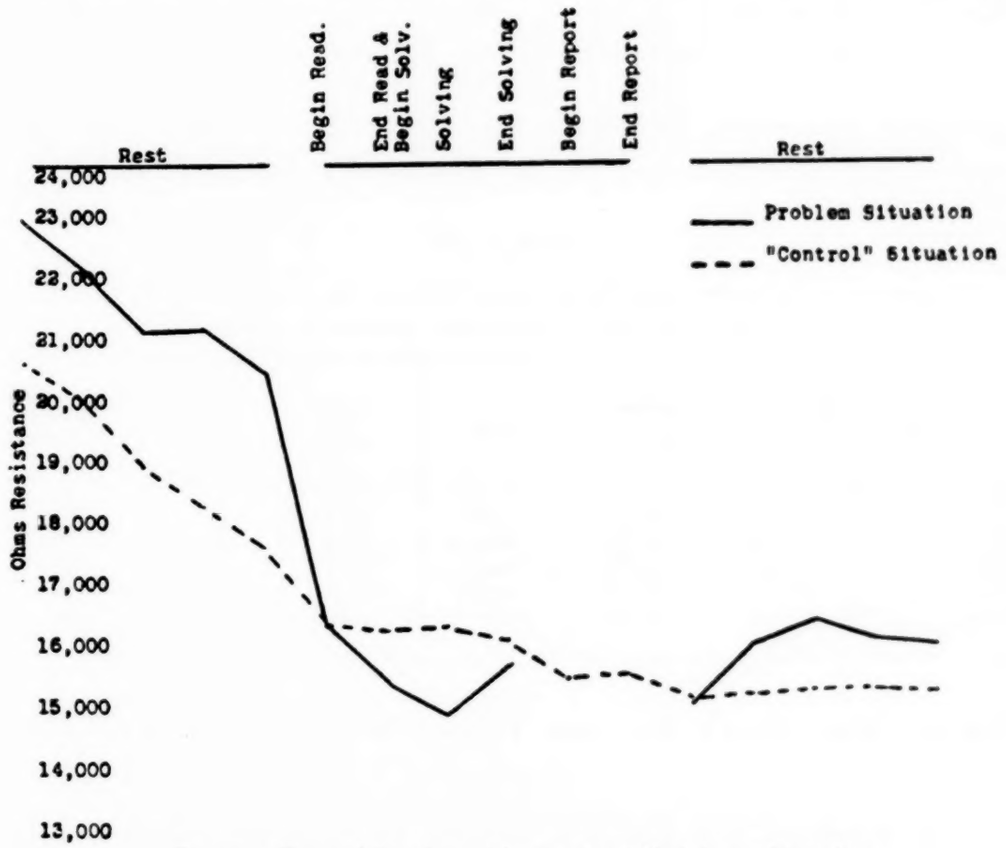

FIG. 2. Record for the subjects who failed (20 subjects)

All but one of the twenty subjects registered a marked decrease at this point.

During the initial rest period there is a consistent tendency for the resistance level in the case of failing subjects to be higher than it is for successful subjects.

When the problem is presented the skin resistance to an electric current in the case of failing subjects is higher than in the case of subjects who succeeded but the difference does not seem to be reliable. During the rest period after the

TABLE III

DIFFERENCE BETWEEN THE RESISTANCE LEVELS FOR SUCCESSFUL SUBJECTS AND FAILING SUBJECTS DURING THE PERIOD OF INITIAL REST

Minutes	Resistance for Failures	Resistance for Successes	Diff.	S.D. of Diff.	$\frac{\text{Diff.}}{\text{S.D.}_{\text{D.}}}$	Chances per 1,000 of Real Difference
First.....	22,925	19,212	3713	1826	2.0	977
Second.....	22,130	18,293	3937	1933	2.0	977
Third.....	21,107	17,786	3321	2114	1.5	934
Fourth.....	21,126	17,010	4116	2151	1.9	971
Fifth.....	20,482	16,478	4004	2278	1.8	964

problem situation, however, this difference tends to become more pronounced as the period passes (see Table IV).

TABLE IV

DIFFERENCE BETWEEN THE RESISTANCE LEVELS FOR SUCCESSFUL SUBJECTS AND FAILING SUBJECTS DURING THE PERIOD OF FINAL REST

Minutes	Resistance for Failures	Resistance for Successes	Diff.	S.D. of Diff.	$\frac{\text{Diff.}}{\text{S.D.}_{\text{D.}}}$	Chances per 1,000 of Real Difference
First.....	14,980	14,060	920	1314	.7	760
Second.....	15,948	13,872	2124	1454	1.5	930
Third.....	16,313	13,863	2450	1527	1.6	940
Fourth.....	16,036	13,620	2416	1611	1.5	930
Fifth.....	15,903	13,746	2257	1520	1.5	930

THOSE WHO GAVE UP AND THOSE WHO PERSEVERED BUT FAILED

An attempt was made to divide the failing subjects into two sections, those who persevered and those who gave up. The criteria used were comments by the subjects concerning the problem, attempted reports, and introspections given by the subjects after the sittings were completed. It is to be admitted that the validity of this classification is questionable and there were but few cases in each division; however, the facts revealed are suggestive of probable distinctions.

In the case of those who gave up there was a very great drop in resistance when they commenced to read the problem. The level even dropped below that of the corresponding control

reading. During the final rest, the resistances are practically the same for the problem and for the control situation.

The resistance level in the case of the persevering subjects in the problem situation tends to remain above that for the control situation. The data here presented do not show this difference to be conventionally reliable but these data and those for the subjects who succeeded show that a real difference exists between the resistance trends for these two groups.

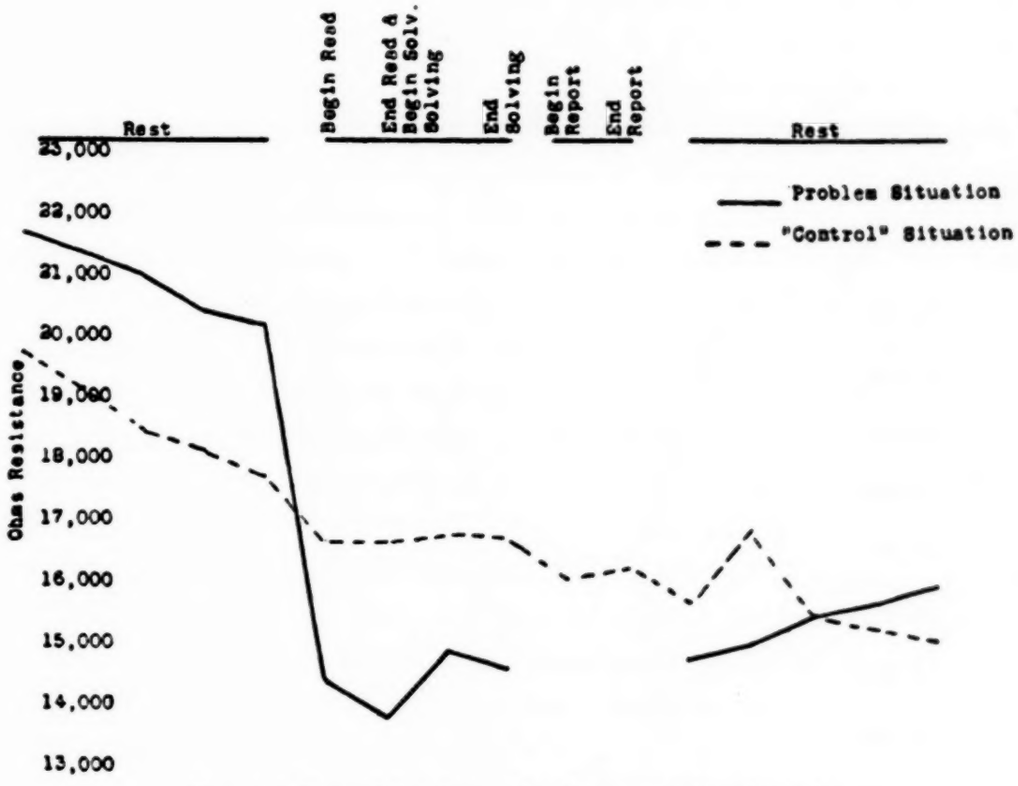

FIG. 3. Record for those who 'Gave up' (8 subjects).

There is less drop in resistance at the beginning of the reading in the persevering than in the giving-up subjects.

CONCLUSIONS

1. In the case of successful subjects there is a very definite tendency for the skin resistance to be less when the problem is about to be attempted than in the similar period when they know there will be no problem to be solved. During the reading of the problem, during solving and during report of the solution the skin resistance is definitely lower than it is for

the same subjects when no problem is presented. This lowered resistance persists for some time subsequent to the reporting of the solution.

Throughout the control situation there is a tendency for the resistance to be higher in the case of successful subjects than in the case of failing subjects.

2. For failing subjects the skin resistance during the initial rest is lower in the control situation than in the prob-

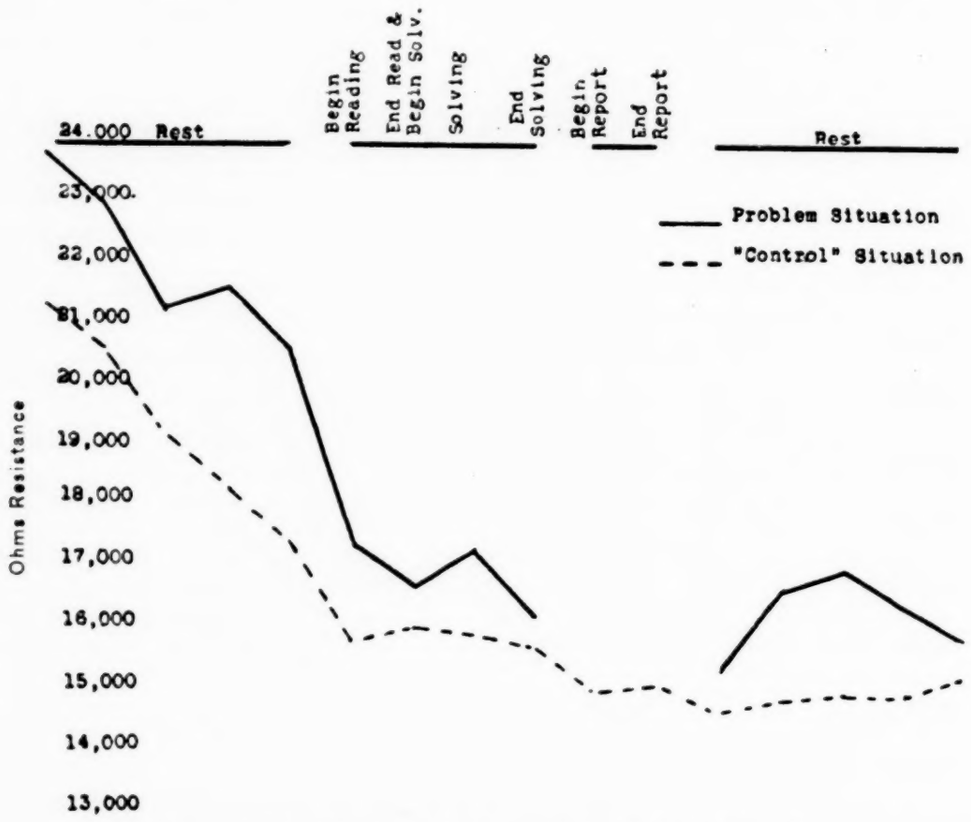

FIG. 4. Record for those who persevered (12 subjects).

lem situation. There is a definite drop in resistance when failing subjects start trying to solve the problem. For these people the resistance is practically the same during attempted solution as it is in the corresponding control.

3. The difference trends between the resistances for successful and failing subjects are real, *i.e.*, not due to sampling error.

4. Subjects who give up have a significant drop in resistance as they start reading the problem. This resistance

tends to rise during attempted solution and the rest period after the problem.

5. The resistance in the case of persevering but failing subjects tends to be higher at all points during the problem situation than it is during the corresponding control situations.

Summary

Successful solving of these problems of the Binet Test is accompanied by decreased skin resistance. This decrease becomes noticeable when a problem is about to be attacked and persists for some time after the successful solution. Persevering but unsuccessful problem solving is accompanied by a noticeable drop in skin resistance to an electric current but the level during attempted solution is not lower than during the control situation. In the case of those who give up the resistance drops significantly upon the presentation of the problem but the resistance tends to rise during attempted solution and during the rest periods after the problem.

Inasmuch as the activity of the sweat glands is under the control of the sympathetic nervous system, a decrease in electrical resistance indicates that during successful problem solving of this type the sympathetic nervous system is more active than during rest periods of these subjects. Unsuccessful problem solving is not accompanied by an increase of activity of the sympathetic nervous system as compared with the activity during rest.

Bibliography

1. Darrow, Chester W., Differences in the physiological reactions to sensory and ideational stimuli, *Psych. Bull.*, **26**, 185–201, 1929.
2. Davis, Roland C., *Ability in social and racial classes.* The Century Company, New York, 1931.
3. —— and Porter, A measuring device for the galvanic reflex, *Jour. Gen. Psych.*, **5**, 115–120, 1931.
4. Collman, Robin D., The psychogalvanic reactions of exceptional and normal children. Teachers College, Columbia University, *Contributions to Education*, no. 469, 1931.
5. Sears, Richard, Psychogalvanic responses in arithmetical work, *Arch. Psych.*, no. 155, 1933.

DISCUSSION

A NOTE ON THE RELATIVE EFFICACY OF A BELL SIGNAL FOR RIGHT AND WRONG RESPONSES IN MAZE LEARNING

BY KARL F. MUENZINGER

The University of Colorado

Hulin and Katz (1) recently repeated part of a maze experiment by Tolman, Hall and Bretnall (3), which had also been repeated by me (2). They concluded that their "results show that the Bell-Right situation was more easily learned . . . than the Bell-Wrong situation. This finding agrees with the . . . results from Tolman, Hall and Bretnall and from Muenzinger." I would have no quarrel with this statement if they had indicated that they referred to only half of my results; then they go on saying that they "consider that this agreement . . . tends to contradict the conclusion [that no significant difference has been demonstrated] proposed by Muenzinger regarding his own data" (1, p. 642). Besides the agreement between their results and part of mine Hulin and Katz offer no reasons to support their claim.

A reading of my article, especially of tables 1 and 2, will show that my repetition and extension of the California experiment was intended to supply comparisons between comparable groups. In repeating the conditions bell-for-right and bell-for-wrong responses I used two parallel sets of two groups each, one set learning pattern A and the other set pattern B of the maze, and one group in each set learning with the bell-right condition and the other with the bell-wrong condition. In the first set the bell-right group was the better one, a result used by Hulin and Katz in pointing out the agreement referred to above. However, *in the second set the bell-wrong group was the better one*, a result not mentioned by Hulin and Katz. In view of opposite results from the two sets of groups I do not see what other conclusion any one could propose except that the results show no significant difference between the bell-right and the bell-wrong conditions.

In repeating the California experiment Hulin and Katz did not use comparable groups, but had one group, bell-right, learn pattern

A and the other group, bell-wrong, learn pattern B of the maze, thus exactly duplicating the California groups II and I respectively. Quite aside from my own insistence on comparable groups, this procedure seems curious when one reads in the article of the California experimenters (3, p. 608) that they themselves had felt the need for comparable groups and added a third group, Ia, learning pattern A with bell-wrong. While the critical ratio (D/P.E. of D) for errors in the case of their non-comparable groups I and II, that is, those duplicated by the Princeton experimenters, had been 1.95 (in the Princeton experiment 2.3), it now dropped down to 0.61 in the case of the comparable groups I and Ia.[1] Even if one chooses to disregard the opposite results from the two sets of comparable groups in my experiment, the small critical ratios involved here would not seem to justify the conclusion supported by Hulin and Katz that a significant difference in favor of the bell-right condition has been demonstrated. The condition bell-right may be more efficacious, but the results from these three experiments do not show it.

References

1. Hulin, W. S., and Katz, D., A comparison of emphasis upon right and upon wrong responses in learning, J. Exper. Psychol., 1935, **18**, 638–642.
2. Muenzinger, K. F., Motivation in learning. II. The function of electric shock for right and wrong responses in human subjects, J. Exper. Psychol., 1934, **17**, 439–448.
3. Tolman, E. C., Hall, C. S., and Bretnall, E. P., A disproof of the law of effect and a substitution of the laws of emphasis, motivation and disruption, J. Exper. Psychol., 1932, **15**, 601–614.

A COMPARISON OF THE BELL SIGNAL FOR RIGHT AND WRONG RESPONSES: A REPLY

BY WILBUR S. HULIN AND DANIEL KATZ

Princeton University

The present writers, on the basis of their own experimental data (1), and that of Tolman, Hall and Bretnall (4), and of Muenzinger (2), concluded that the sounding of a bell accompanying right responses aided learning more than the sounding of a bell

[1] Hulin and Katz are mistaken in stating that both California groups I and II learned pattern A (1, p. 441, note 1); they learned patterns B and A respectively Another similar error occurs on the same page (1, p. 441, note 2).

accompanying wrong responses. Muenzinger (3) has raised two objections to the conclusions of the writers.

In the first place Muenzinger claims that the writers have considered only part of his results which are pertinent to the problem. He states, "In the first set the bell-right group was the better one, a result used by Hulin and Katz in pointing out the agreement referred to above. However, *in the second set the bell-wrong group was the better one*, a result not mentioned by Hulin and Katz" (3).

In this second set the average error for the bell-wrong group for twenty trials is only 0.9 less than for the bell-right group (BR 125.5–BW 124.6) (2, p. 441); the critical ratio is 0.2. This is the basis for Muenzinger's statement that "*the bell-wrong group was the better one.*" In contrast to this difference of 0.9 are the differences obtained by Muenzinger in his other set of experiments, by Tolman, Hall and Bretnall, and by the present writers. Muenzinger's first set of experiments showed a difference in average errors in favor of the bell-right group of 29.1, critical ratio 3.3 (2, p. 441); Tolman, Hall and Bretnall report a difference in favor of the bell-right group of 13.41, critical ratio 1.95 (4, p. 607); and the present writers found a difference of 17.8, critical ratio 2.3, in favor of the bell-right situation (1, p. 642). Thus in the four experimental comparisons the bell-right situation has been more efficacious for learning in three cases. Hence the writers still think the evidence favors the bell-right situation over the bell-wrong situation, especially since the experiments were carried out in three different laboratories.

Muenzinger's second objection is that the writers did not control the factor of the pattern learned by the subjects. The assumption is that differences in learning might be due to a difference in pattern. The present writers did call attention to the results which Muenzinger and Tolman, Hall and Bretnall obtained on this question and concluded that "no consistent difference between the two patterns was shown" (1, p. 639). This agrees with Muenzinger's own conclusion: "If we compare the difficulty of pattern A with that of pattern B, we again get contradictory results so that we cannot assume a reliable difference between the two situations" (2, p. 443).

References

1. Hulin, W. S., and Katz, D., A comparison of emphasis upon right and upon wrong responses in learning, J. Exper. Psychol., 1935, **18**, 638–642.

2. MUENZINGER, K. F., Motivation in learning. II. The function of electric shock for right and wrong responses in human subjects, J. EXPER. PSYCHOL., 1934, **17**, 439–448.
3. *idem*, A note on the relative efficacy of a bell signal for right and wrong responses in maze learning, J. EXPER. PSYCHOL., 1936, **19**.
4. TOLMAN, E. C., HALL, C. S., AND BRETNALL, E. P., A disproof of the law of effect and a substitution of the laws of emphasis, motivation and disruption, J. EXPER. PSYCHOL., 1932, **15**, 601–614.